Review — Independent Trials

- Independent trials: the outcomes of the various trials do not influence or affect one another in any way
- Independence of trials is a belief and cannot be proved mathematically
- Compound experiment = independent trials of a simple experiment
- Simple versus compound events
- $P(A, B, C, A^c, \ldots) = P(A)P(B)P(C)P(A^c)\ldots$

Review — Random Variables

- **X** is a random variable defined on the simple experiment.
- $\mathbf{X}_i$ is the value of **X** on i-th subexperiment
- $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \ldots)$ is called a random vector
- For repeated independent trials, the random variables $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4, \ldots$ are said to be independent random variables
  $P(\mathbf{X}_1 = a_5, \mathbf{X}_2 = a_2, \mathbf{X}_3 = a_7, \mathbf{X}_4 = a_9, \ldots) =$
  $= P(\mathbf{X}_1=a_5)P(\mathbf{X}_2=a_2)P(\mathbf{X}_3=a_7)P(\mathbf{X}_4=a_9)\ldots$

Binomial Random Variables

- A binomial random variable **Y** with parameters (n, p) is the number of times an event A of probability p occurs on n independent trials
- **Y** takes on values 0, 1, 2, … n
- For $0 \le k \le n$,
  $p_Y(k) = P\{\mathbf{Y} = k\} = \binom{n}{k} p^k(1-p)^{n-k}$
- P(A occurred on a specific set of k trials) $= p^k(1-p)^{n-k}$; $P\{\mathbf{Y} = k\}$ is the probability that A occurred on some set of k trials

Probabilities from the binomial pmf

- Probabilities such as $P\{a < \mathbf{Y} < b\}$ are found by summing up the appropriate terms in the pmf
- There are no closed-form expressions for such probabilities: numerical evaluation is required
- $P\{a < \mathbf{Y} < b\}$ is the sum of L (say) of the n+1 probabilities in the pmf
- If L > (n+1)/2, find 1 – P{complement}

Mean of binomial random variable

- $E[\mathbf{Y}] = np$
- $E[\mathbf{Y}] = \sum_{k=0}^{n} k \bullet \binom{n}{k} p^k(1-p)^{n-k}$
  $= \sum_{k=1}^{n} k \bullet \binom{n}{k} p^k(1-p)^{n-k}$
  $= np \bullet \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1}(1-p)^{(n-1)-(k-1)}$
- $E[\mathbf{Y}] = np \bullet (p + 1 - p)^{n-1} = np$

Variance of binomial RV

- $\mathrm{var}(\mathbf{Y}) = np(1-p)$
- $E[\mathbf{Y}(\mathbf{Y}-1)] = E[\mathbf{Y}^2] - E[\mathbf{Y}]$
- $E[\mathbf{Y}(\mathbf{Y}-1)] = \sum_{k=0}^{n} k(k-1) \bullet \binom{n}{k} p^k(1-p)^{n-k}$
  $= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2}(1-p)^{(n-2)-(k-2)}$
- $E[\mathbf{Y}(\mathbf{Y}-1)] = n(n-1)p^2 \bullet (p + 1 - p)^{n-2}$
  $= n(n-1)p^2$

Variance of binomial RV– continued

- $E[\mathbf{Y}(\mathbf{Y}-1)] = E[\mathbf{Y}^2] - E[\mathbf{Y}] = n(n-1)p^2$
- $E[\mathbf{Y}] = np$
- $\text{var}(\mathbf{Y}) = E[\mathbf{Y}^2] - (E[\mathbf{Y}])^2$
  $= E[\mathbf{Y}(\mathbf{Y}-1)] + E[\mathbf{Y}] - (E[\mathbf{Y}])^2$
  $= n(n-1)p^2 + np - n^2p^2$
  $= np - np^2 = np(1-p)$
- $\text{var}(\mathbf{Y}) = np(1-p)$
- Maximum value of var(**Y**) is $n/4$ at $p = 1/2$; else $\text{var}(\mathbf{Y}) < n/4$; $\text{var}(\mathbf{Y}) = 0$ if $p = 0$ or 1

Statistical Estimation—Introduction

- For a binomial random variable **Y** with parameters (n, p), the question "What is the value of P{**Y** = k}?" has a definitive and precise answer $\binom{n}{k} p^k(1-p)^{n-k}$
- We now consider a different question
- Suppose that p, the probability of the event A is unknown
- Given that A occurred k times on n trials, that is, we observed the event {**Y** = k}, what is a good estimate of the value of p?

Estimation of probabilities

- Our paradigm here is of a coin of unknown provenance
- In order to estimate p = P(Heads) for this coin, we toss it n times (independent trials!) and observe k Heads on the n tosses
- What should we estimate p to be?
- Relative frequency estimate is
  $$\hat{p} = k/n$$
- The estimate is $\hat{p}$; the actual value is p

Point Estimates

- Relative frequency estimate of p is k/n
- This is called a point estimate of p: the estimated value of p is a number or point in the interval [0, 1]
- Example: 5,023 Heads are observed on 10,000 tosses. The relative frequency estimate of p is 0.5023
- Is it possible that actually p = 0.5 instead?
- Is it possible that actually p = 0.51?

Interval Estimates

- An interval estimate estimates the value of p as being in an interval (a, b) or [a, b]
- Example: 5,023 Heads are observed on 10,000 tosses. An interval estimate is of the form
  "$0.4973 < p < 0.5073$"
  "$0.5013 \le p \le 0.5033$"
- The length of the interval is a crucial parameter of the estimate

Confidence levels

- How sure are we that the unknown value of p actually is in the interval specified?
- If we specify the interval as [0, 1], then we can be 100% confident that p lies in the specified interval
- For smaller intervals, we naturally have a lesser degree of confidence
- We are more sure about the assertion that "$0.4973 < p < 0.5073$" than about the assertion that "$0.5013 \le p \le 0.5033$"

Confidence intervals and levels

- The interval that we specify is usually called a confidence interval; associated with it is a confidence level
- Confidence level measures the degree of our belief that the actual value of p does lie in the confidence interval
- In reports on polls, the “margin of error” tells the length of the confidence interval; the confidence level is usually unspecified
- Wide interval ⇔ high confidence level

Remarks on Estimation

- We are asked for an estimate of p
  - point estimates based on various methodologies and reasonings
  - interval estimates of various lengths and levels of confidence
  - ignore the data; always estimate p = 0.5
- P{**Y** = k} has precise unambiguous answer No single statistical estimate is “the one and only correct estimate” in contrast with which all others are godless inventions

Maximum-Likelihood Principle I

- We have observed an event A of unknown probability p occurring k times on n trials
- We need to estimate the value of p
- If our estimate is the number p_1, then we are claiming that we have just observed an event of probability $\binom{n}{k}p_1^k(1-p_1)^{n-k}$
- If our estimate is the number p_2, then we are claiming that we have just observed an event of probability $\binom{n}{k}p_2^k(1-p_2)^{n-k}$

Maximum-Likelihood Principle II

- Should we use p_1 or p_2 as our estimate of p? Which is more reasonable?
- Fundamental notion: It is more reasonable to assume that a probable event occurred rather than an improbable one
- We should compare the two probabilities in question and estimate p as that number which gives a larger probability for the event that we just observed:
  A occurred k times on n trials

Maximum-Likelihood Principle III

- Statisticians call the probability
  $$\binom{n}{k}p_1^k(1-p_1)^{n-k}$$
  the likelihood of the observed event (k occurrences of A on n trials) under the assumption that $p = p_1$)
- Why not just call it probability?
- The different likelihoods obtained under the different assumptions about p do not add up to 1 (like probabilities should!)

Maximum-Likelihood Principle IV

- We are asked for an estimate of p
- Maximum-likelihood principle: the estimate should be the number that maximizes the likelihood of the observation
- The likelihood of the observation under the assumption that P(A) = x is just
  $$\binom{n}{k}x^k(1-x)^{n-k}$$
- What value of x, $0 \le x \le 1$, maximizes this function?

Maximum-Likelihood Principle V

- What value of x, $0 \le x \le 1$, maximizes the function $x^k(1-x)^{n-k}$?
- n-choose-k is just a constant that does not affect the location of the maximum
- The function $x^k(1-x)^{n-k}$ is 0 at x = 0 and at x = 1; it is positive for 0 < x < 1
- The derivative of $x^k(1-x)^{n-k}$ is 0 at x = k/n
- Exercise: Work through the details and check this result

Maximum-Likelihood Principle VI

- The derivative of $x^k(1-x)^{n-k}$ is 0 at x = k/n
- $x^k(1-x)^{n-k}$ has a maximum at x = k/n
- Exercise: Why can we assert that x = k/n is a maximum (not a minimum) without checking whether the second derivative is negative?
- Summary: If event A of unknown probability p occurred k times on n independent trials, then the maximum likelihood (ML) estimate of p is k/n

Justification of relative frequency

- Event A occurred k times on n trials
- The relative frequency estimate (a point estimate) of P(A) = p is k/n
- This is so natural an estimate that even beginners in probability theory accept it
- It is also an estimate of great antiquity
- It is comforting to know that the relative frequency estimate of p can be justified as the maximum-likelihood estimate of p

A closer look at the data?

- We have discussed the estimation of the unknown probability in the case when event A has occurred k times on n trials
- Could we have extracted a better estimate of p by considering the actual sequence of occurrences of A and A^c?
- No
- $P(A, A, A^c, A, \ldots) = P(A)P(A)P(A^c)P(A)\ldots = p^k(1-k)^{n-k}$ so we get the same result!

A critique of the method I

- If we use n trials of the experiment, the maximum-likelihood estimate of p will necessarily be one of the numbers 0/n, 1/n, 2/n, 3/n, … , n/n
- To get a more precise (though not necessarily more accurate!) estimate of p, we need to increase the number of trials
- But what is the accuracy of the maximum-likelihood estimate anyway?

A critique of the method II

- The value of the unknown parameter p is assumed to be any number in [0, 1]
- It is only in extremely fortuitous circumstances that an ML estimate will be the exact value of p
- More often than not, the ML estimate will be close, but not exactly right
- Attempting to get a more precise estimate only reduces the chances of accuracy

A critique of the method III

- Example: The pmf of a binomial random variable with parameters (10, 0.5) is shown in Ross, p. 151
- The ML estimate will be 0.5 (the right answer) if and only if the event {**Y** = 5} occurs, which has probability less than 0.25 of occurring
- Put another way,

  P{ML estimate ≠ 0.5} > 0.75

A critique of the method IV

- To get two decimal digits of precision in the estimate of p, we need n = 100 trials
- If **Y** is a binomial random variable with parameters (100, 0.5), then

  P{**Y** = 50} < 0.08
- The ML estimate will be 0.5 (the right answer) only if {**Y** = 50} is observed
- Put another way,

  P{ML estimate ≠ 0.5} > 0.92

The mode of the binomial pmf I

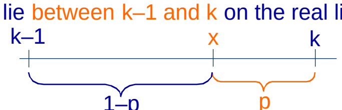
- The value of u that maximizes $p_X(u)$ is called the mode of the random variable **X**
- What choice of k maximizes the probability that event A of known probability p will occur k times on n trials?
- Put another way, if you have to bet on the value that a binomial random variable with parameters (n, p) will take on, which of the numbers 0, 1, 2, … , n should you choose?

The mode of the binomial pmf II

- Ross (Proposition 7.1, p. 150) shows that for each choice of k, 1 ≤ k ≤ n
  - P{**Y** = k} > P{**Y** = k–1} if k < (n+1)p
  - P{**Y** = k} = P{**Y** = k–1} if k = (n+1)p
  - P{**Y** = k} < P{**Y** = k–1} if k > (n+1)p
- The maximum value of p(k) = P{**Y** = k} is at k = $\lfloor (n+1)p \rfloor$, nearly the mean value np
- It is easier to understand this result graphically as shown on the next slide

A picture is worth a thousand words…

- For some number k, 1 ≤ k ≤ n, the point np will lie between k–1 and k on the real line



- The pmf values increase from 0 upto k–1 The pmf values decrease from k to n
- Max pmf value is at k–1 or k according as point np is to the left or right of the point x
- If np = x, p(k–1) = p(k) are twin peaks

Maximum-Likelihood Principle VII

- Given that an event A of known probability occurred k times on an unknown number n of trials, we can find the ML estimate of n
- Example: A student has marked the wrong answer to k multiple-choice questions. Each question has 5 choices. The student guesses at random on questions whose answers he does not know, and gets the wrong answer with probability 0.8. How many questions was he guessing on?

Maximum-Likelihood Principle VIII

- The ratio method of Ross's Proposition 7.1 of Chapter 4 can be applied here
- Let b(k; n, p) denote the probability that a binomial random variable with parameters (n, p) has value k
- k and p are the known quantities
- We look at the ratio b(k; n, p)/b(k; n–1, p) = n(1–p)/(n–k)
- The ML estimate of n is ≈ k/p

Maximum-Likelihood Principle IX

- In Section 4.9.3, Ross solves a similar problem for a hypergeometric random variable
- Hypergeometric random variables arise in sampling without replacement
- A sample of size n is chosen from a set of size N that contains m items of one kind and N–m items of another kind
- The sample contains i items of first kind

Maximum-Likelihood Principle X

- Given that i items in the sample of size n are of the first kind, what is ML estimate of m assuming that N is known? ≈ iN/n
- Applications to quality control and acceptance sampling
  - If i items in the sample are defective, how many defectives in the whole set?
  - Should the whole set be shipped?
- Relative frequencies again: i/n ≈ m/N

Maximum-Likelihood Principle XI

- Given that i items in the sample of size n are of first kind, what is the ML estimate of N assuming that m is known? ≈ mn/i
- Applications to biology and ecology
  - Estimate the number of fishes in the sea by introducing m "tagged" specimens and then determine how many tagged fish are caught on a later fishing expedition that catches n fishes
- Relative frequencies again: m/N ≈ i/n

Summary

- Basic properties of binomial random variables with parameters (n, p)
- E[**Y**] = np   var(**Y**) = np(1–p) ≤ n/4
- Most likely value (mode) of **Y** is $\lfloor (n+1)p \rfloor$
- Point estimates versus interval estimates
- Confidence levels and confidence intervals
- The narrower the confidence interval, the lower the confidence level

Summary (continued)

- Principle of maximum-likelihood: the estimate of the value of an unknown parameter is the number that maximizes the likelihood of the observation
- If an event A occurred k times on n trials, the maximum-likelihood estimate of P(A) is k/n, the observed relative frequency of A
- Maximum likelihood estimation of discrete parameters