Review I

- **Y** denotes the number of occurrences of an event A of probability p on n trials
- **Y** is a binomial random variable with parameters (n, p). It has mean E[**Y**] = np, variance np(1–p), and mode $\lfloor (n+1)p \rfloor$
- Problem: **Y** had value k on a trial of the (compound) experiment. Estimate the unknown value of p from this datum
- The relative frequency estimate of p is k/n

Review II

- The relative frequency estimate can be justified via maximum-likelihood principle
- Maximum-likelihood (ML) principle: 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 ML estimate of P(A) is k/n, which is the same as the relative frequency estimate

Review III

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

Review IV

- Point estimates (e.g. the ML estimate k/n) versus interval estimates such as

  "0.5013 < p < 0.5033"

  or

  "k/n ± 3%"
- Associated with a confidence interval is its confidence level
- The narrower the confidence interval, the lower the confidence level

Confidence intervals and levels

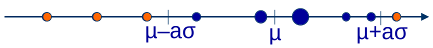
- How does one find a confidence interval?
- If 5,023 Heads occurred on 10,000 tosses of a coin, and we want a confidence interval of length 0.1, where in the interval [0,1] should our confidence interval be?
- What is the confidence level associated with this confidence interval?
- How to find a confidence interval for a specified confidence level?

Back to probability … for a while

- The variance of a random variable is a measure of the spread of the probability masses about the mean µ
- The larger the variance, the wider the dispersion of the masses away from µ
- How much probability mass lies "far away" from µ?
- How much probability mass is at distance aσ or more from µ?

Re: being way out in left field …

- How much probability mass is at distance $a\sigma$ or more from μ?



- $P\{X \geq \mu + a\sigma\} + P\{X \leq \mu - a\sigma\}$
  $= P\{|X - \mu| \geq a\sigma\}$
  = total masses in orange
- $P\{|X - \mu| \geq a\sigma\} = \sum_{u_i \geq \mu + a\sigma} p(u_i) + \sum_{u_j \leq \mu - a\sigma} p(u_j)$

So, what's so hard about that?

- For any given pmf, $P\{|X - \mu| \geq a\sigma\}$ can be calculated straightforwardly
- However, there is a more generic answer to the question
  "How much probability mass is at distance $a\sigma$ or more from μ?
- No matter what the pmf happens to be,
  $P\{|X - \mu| \geq a\sigma\} \leq 1/a^2$
- This is called the Chebyshev inequality

The Chebyshev Inequality

- Chebyshev inequality: For any random variable with mean μ and variance σ^2
  $P\{|X - \mu| \geq a\sigma\} \leq 1/a^2$
- This result also applies to continuous random variables (studied later)
- No more than $1/a^2$ of the probability mass is at distance of $a\sigma$ or more from the mean
- Usually, far less than $1/a^2$ mass is so far away from the mean

Proof of Chebyshev Inequality I

- Chebyshev inequality: For any random variable with mean μ and variance σ^2
  $P\{|X - \mu| \geq a\sigma\} \leq 1/a^2$
- $P\{|X - \mu| \geq a\sigma\} = \sum_{u_i \geq \mu + a\sigma} p(u_i) + \sum_{u_j \leq \mu - a\sigma} p(u_j)$
- $(u_i - \mu)/a\sigma \geq 1$ and $(u_j - \mu)/a\sigma \leq -1$
- We make each term on the right larger by
  - multiplying $p(u_i)$ by $[(u_i - \mu)/a\sigma]^2 \geq 1$
  - multiplying $p(u_j)$ by $[(u_j - \mu)/a\sigma]^2 \geq 1$

Proof of Chebyshev Inequality II

- $P\{|X - \mu| \geq a\sigma\} = \sum_{u_i \geq \mu + a\sigma} p(u_i) + \sum_{u_j \leq \mu - a\sigma} p(u_j)$
  $\leq \sum_{u_i \geq \mu + a\sigma} [(u_i - \mu)/a\sigma]^2 p(u_i) + \sum_{u_j \leq \mu - a\sigma} [(u_j - \mu)/a\sigma]^2 p(u_j)$
- Next, we add on the right side the positive term $[(u_i - \mu)/a\sigma]^2 p(u_i)$ for each u_i that satisfies $\mu - a\sigma < u_i < \mu - a\sigma$, thus further increasing the sum
- Sum is now over all values taken on by **X**

Proof of Chebyshev Inequality III

- $P\{|X - \mu| \geq a\sigma\} \leq \sum_{\text{all } u_i} [(u_i - \mu)/a\sigma]^2 p(u_i)$
  $= [a\sigma]^{-2} \sum_{\text{all } u_i} [(u_i - \mu)]^2 p(u_i)$
- But, by definition, the sum is just σ^2
- Hence, we have just shown that
  $P\{|X - \mu| \geq a\sigma\} \leq 1/a^2$
  which is the Chebyshev inequality

Some obvious numerology

- $P\{|\mathbf{X} - \mu| \geq a\sigma\} \leq 1/a^2$
- Note that we must choose $a > 1$ for this inequality to be of any use
- No more than 25% of the probability mass is at distance 2σ or more from μ
- No more than 4% of the probability mass is at distance 5σ or more from μ
- No more than 1% of the probability mass is at distance 10σ or more from μ

The Chebyshev Inequality is weak

- Consider a binomial random variable **X** with parameters (10, 0.5)
- $\mu = 5$, $\sigma^2 = 2.5$, $\sigma = 1.58\ldots$
- Chebyshev Inequality gives
  $P\{|\mathbf{X} - 5| \geq 3.16\} \leq 0.25$
- Actually, $P\{|\mathbf{X} - 5| \geq 3.16\}$
  $= P\{\mathbf{X} \leq 1.84\} + P\{\mathbf{X} \geq 8.16\}$
  $= P\{\mathbf{X} = 0\}+P\{\mathbf{X} = 1\}+P\{\mathbf{X} = 9\}+P\{\mathbf{X} = 10\}$
  $= 22/1024 = 0.0215\ldots \ll 0.25$

What's a lousy inequality good for?

- The importance of the Chebyshev inequality lies in its generality — it can be applied in a wide variety of cases
- The Chebyshev inequality allows us to state general results that apply to all random variables
- Example: For any random variable, at least 96% of the probability mass lies in the interval $(\mu - 5\sigma, \mu+5\sigma)$

Applying the Chebyshev inequality

- Let **X** denote a binomial random variable with parameters (n, p)
- On a trial of the (compound) experiment, we observe that **X** had value k, $0 \leq k \leq n$
- The probability is at least 96% that the observed value k is in the interval
  $\left(np - 5[np(1-p)]^{1/2}, np + 5[np(1-p)]^{1/2}\right)$

Changing the assumptions a bit …

- **X** is a binomial (n, p) random variable
- p is an unknown parameter
- On a trial, the event $\{\mathbf{X} = k\}$ occurred
- The probability is at least 96% that the observed value k is in the interval
  $\left(np - 5[np(1-p)]^{1/2}, np + 5[np(1-p)]^{1/2}\right)$
- We don't know what p is, but we do know (we do?) that $np(1-p) \leq n/4$

Bigger interval includes smaller …

- **X** is a binomial (n, p) random variable with unknown parameter p
- Event $\{\mathbf{X} = k\}$ occurred
- The interval
  $\left(np - 5[np(1-p)]^{1/2}, np + 5[np(1-p)]^{1/2}\right)$
  is a subset of the interval
  $\left(np - 5[n/4]^{1/2}, np + 5[n/4]^{1/2}\right)$
  i.e. $(np - 2.5\sqrt{n}, np + 2.5\sqrt{n})$

P{X   bigger} ≥ P{X   smaller}

- **X** is a binomial (n, p) random variable with unknown parameter p
- On a trial, the event {**X** = k} occurred
- The probability is at least 96% that the observed value k is in the interval

$$\left(np - 5[np(1-p)]^{1/2},\ np + 5[np(1-p)]^{1/2}\right)$$

- The probability is at least 96% that the observed value k is in the interval

$$(np - 2.5\sqrt{n},\ np + 2.5\sqrt{n})$$

Estimate of p is in interval …

- **X** is a binomial (n, p) random variable with unknown parameter p. {**X** = k} occurred
- Maximum likelihood estimate of p is k/n
- The probability is at least 96% that the observed value k is in the interval

$$(np - 2.5\sqrt{n},\ np + 2.5\sqrt{n})$$

- The probability is at least 96% that the ML estimate of p is in the interval

$$(p - 2.5/\sqrt{n},\ p + 2.5/\sqrt{n})$$

A really stupid statement?

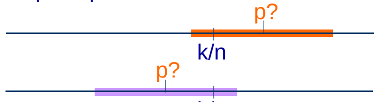
- **X** is a binomial (n, p) random variable with unknown parameter p. {**X** = k} occurred
- The probability is at least 96% that k/n, the ML estimate of p, is in the interval

$$(p - 2.5/\sqrt{n},\ p + 2.5/\sqrt{n})$$

- The length of the interval is $5/\sqrt{n}$, but the mid-point is the unknown quantity p
- What good does this do us?
- We can try and fudge the data!

Where's Waldo? Where's p?

- **X** is a binomial (n, p) random variable with unknown parameter p. {**X** = k} occurred
- Maximum likelihood estimate of p is k/n
- With at least 96% probability, k/n is in the interval $(p - 2.5/\sqrt{n},\ p + 2.5/\sqrt{n})$ whose midpoint p is unknown



Arguing bass ackwards…

- With at least 96% probability, k/n is in the interval $(p - 2.5/\sqrt{n},\ p + 2.5/\sqrt{n})$ whose midpoint p is unknown
- But if this event (viz. **X** is in the interval $(np - 2.5\sqrt{n},\ np + 2.5\sqrt{n})$) of probability at least 0.96 occurred, then the ML estimate k/n of p is within $\pm 2.5/\sqrt{n}$ of p
- But this means that with "probability at least 0.96", p is in the interval $k/n \pm 2.5/\sqrt{n}$!

The confidence trick

- But this means that with "probability at least 0.96", p is in the interval $k/n \pm 2.5/\sqrt{n}$!
- Why is the phrase probability at least 0.96 in quotation marks?
- Because there is no probability involved here in the "usual" sense of the word!
- But didn't it all sound perfectly reasonable when it was just explained?
- What's all this no probability business?

How to fool all the people all the time…

- **X** is a binomial (n, p) random variable with unknown parameter p. {**X** = k} occurred
- Consider the interval

  $(k/n - 2.5/\sqrt{n},\ k/n + 2.5/\sqrt{n})$
- Either the parameter p is in this interval, or it is not
- We don't know which of these two states of affairs holds for our data, but clearly, probability has nothing to do with it

Nomenclature

- Statisticians call the interval

  $(k/n - 2.5/\sqrt{n},\ k/n + 2.5/\sqrt{n})$

  or $k/n \pm 2.5/\sqrt{n}$ a confidence interval
- They call the "0.96 probability" as a confidence level rather than a probability because it measures the degree of belief that the unknown parameter p is, in fact, in this confidence interval
- I am 96% confident that p is in the interval

Philosophical dilemmas

- Whether probability is a measure of the degree of belief is a matter that has been argued over by philosophers of probability
- Some believe that all probabilities are subjective, and are just a quantification of the degree of belief
- "I think the probability is 95% that the Earl of Oxford wrote Shakespeare's plays"
- Is this a probability statement?

Back to basics — the original B2B

- **X** is a binomial (n, p) random variable with unknown parameter p. {**X** = k} occurred
- The interval

  $(k/n - 2.5/\sqrt{n},\ k/n + 2.5/\sqrt{n})$

  is called a confidence interval for p
- The confidence level associated with this interval is 96%
- We are 96% confident that p does lie in this confidence interval

More generally …

- **X** is a binomial (n, p) random variable with unknown parameter p. {**X** = k} occurred
- The interval

  $(k/n - a/2\sqrt{n},\ k/n + a/2\sqrt{n})$

  is a confidence interval for p and the associated confidence level is $1 - 1/a^2$
- If we increase a, the length of the confidence interval increases, but so does the confidence level

Trade-offs I

- $(k/n - a/2\sqrt{n},\ k/n + a/2\sqrt{n})$ is a confidence interval for the probability p at confidence level $1 - 1/a^2$
- If a confidence interval of length L is desired, the corresponding confidence level is $C = 1 - 1/nL^2$
- If a confidence level C is desired, the length of the confidence interval is $L = 1/[(n(1 - C)]^{1/2}$

Trade-offs II

- $(k/n - a/2\sqrt{n}, k/n + a/2\sqrt{n})$ is a confidence interval for p at confidence level $1 - 1/a^2$
- The wider the confidence interval, the higher the confidence level
- The narrower the confidence interval, the lower the confidence level
- How can we get a narrow confidence interval with a high confidence level?
- Increase n !!

Trade-offs III

- We can get a very narrow confidence interval with a high confidence level by increasing n, the number of trials of the subexperiment
- In practical situations, costs of the repeated trials must be taken into account and balanced against the costs of obtaining a wrong estimate of p

Trade-offs IV

- Consider estimating the percentage of defective widgets in a production run
- Price of widget takes into account the expected number of defectives, warranties return and replacement policies, etc
- Testing is expensive
- But underestimating p can also have serious economic (and job security!) consequences

How good is all this stuff?

- We began with a weak probability bound
- We used the maximum possible value of variance and did not adjust for $p \neq 0.5$
- By using other results, it is possible to improve matters
- For a given confidence level, the interval can be made narrower than described in this lecture
- But the interval is still of length $O(\sqrt{n})$

How much variation to expect?

- A fair coin is tossed10,000 times. Heads should come up roughly 5,000 times
- Mean = 5,000; variance = 2500; $\sigma = 50$
- The Chebyshev inequality says that 99% of the time we should expect to see between 4,500 and 5,500 Heads ($\pm 10\sigma$)
- The normal approximation says that more than 99% of the time, we will see between 4850 and 5150 Heads ($\pm 3\sigma$)

Is this a biased coin?

- A coin is tossed 10,000 times
- If the coin is fair, then with probability more than 99%, between 4850 and 5150 Heads will be observed
- If the number of Heads observed is not in this range, the coin probably is biased
- But this does not prove the coin is biased; it might be fair, and we were just unlucky enough to get a <1% probability event

Biased beyond a reasonable doubt

- A coin is tossed 10,000 times
- If the number of Heads observed is not in the range 4850-5150, the coin most likely is biased
- It is reasonable to decide that the coin is biased
- It is possible that this decision is wrong!
- In fact, it is wrong with probability less than 1%. Others may demand less than 0.1%?

How say you? Guilty or Not Guilty?

- A coin is tossed 10,000 times
- The number of Heads observed is in the range 4850-5150
- Is it reasonable to assert that the coin is indeed a fair coin?
- No, it might have P(Heads) = 0.5000001
- It is only reasonable to say that there is not sufficient evidence to call it biased
- The verdict is Not Guilty; never Innocent

Decisions, decisions, decisions…

- A coin is tossed 10,000 times
- If the number of Heads observed is not in the range 4850-5150, it is reasonable to assume that the coin is a biased coin
- If the number of Heads observed is in the range 4850-5150, there is not sufficient reason to declare the coin to be biased
- We do not ever say "Yes, the coin is fair"

Summary

- We stated and proved the Chebyshev inequality
- We studied the concepts of confidence interval and confidence level
- We used the Chebyshev inequality to relate the width of the interval and the confidence level
- We discussed the concept of trial by jury in Anglo-Saxon jurisprudence