Math 425
Introduction to Probability
Lecture 36

Kenneth Harris
kaharri@umich.edu

Department of Mathematics
University of Michigan

April 18, 2009

Chebyshev's Inequality

☞ A measure of the concentration of a random variable X near its mean μ is its variance σ^2.

☞ Chebyshev's Inequality says that the probability that X lies outside an arbitrary interval $(\mu - \varepsilon, \mu + \varepsilon)$ is negligible, provided the ratio $\frac{\sigma}{\varepsilon}$ is sufficiently small.

Proposition (Chebyshev's Inequality)

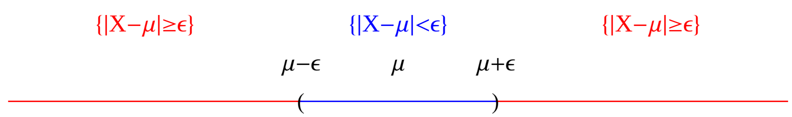
If X is a random variable with finite mean μ and variance σ^2, then for any value $\varepsilon > 0$,

$$\mathbf{P}\{\,|X-\mu| \geq \varepsilon\,\} \leq \frac{\sigma^2}{\varepsilon^2}.$$

Proof of Chebyshev's Inequality

☞ Fix $\varepsilon > 0$. For clarity we take X to be continuous with density $f(x)$.

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\,dx &= \int_{\mu-\varepsilon}^{\mu+\varepsilon} f(x)\,dx + \int_{|x-\mu|\geq\varepsilon} f(x)\,dx \\ &\geq \int_{|x-\mu|\geq\varepsilon} f(x)\,dx \\ &= \mathbf{P}\{\,|X-\mu| \geq \varepsilon\,\}.\end{aligned}$$



Proof of Chebyshev's Inequality

☞ Compute the variance $Var(X) = \sigma^2$:

$$\begin{aligned}\sigma^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\,dx \\ &\geq \int_{|x-\mu|\geq\varepsilon} (x-\mu)^2 \cdot f(x)\,dx \\ &\geq \varepsilon^2 \int_{|x-\mu|\geq\varepsilon} f(x)\,dx \qquad (x-\mu)^2 \geq \varepsilon^2 \\ &= \varepsilon^2 \cdot \mathbf{P}\{\,|X-\mu| \geq \varepsilon\,\}.\end{aligned}$$

So indeed,

$$\mathbf{P}\{\,|X-\mu| \geq \varepsilon\,\} \leq \frac{\sigma^2}{\varepsilon^2}.$$

Note. The proof for discrete X is exactly the same but for replacing "$\int$" with "$\sum$" and density "$f(x)$" with mass "$p(x)$".

Example

Example. Chebyshev's inequality is the best possible inequality – in that there are random variables for which the inequality is in fact an equality.

☞ Fix $\varepsilon > 0$ and choose X with distribution

$$p_X(-\varepsilon) = \frac{1}{2} \qquad p_X(\varepsilon) = \frac{1}{2}$$

So,

$$\mu = E[X] = 0 \qquad \sigma^2 = Var(X) = \varepsilon^2,$$

and thus,

$$1 \;=\; \mathbf{P}\left\{|X - \mu| \geq \varepsilon\right\} \;\leq\; \frac{\sigma^2}{\varepsilon^2} \;=\; 1.$$

Note: this distribution cuts-out the center $(\mu - \varepsilon, \mu + \varepsilon)$ and places the entire mass of the tail of the distribution on $\pm\sigma$.

Example

☞ Chebyshev's inequality is quite crude, and we will see examples where it provides a poor bound. However

1. Chebyshev's inequality provides a bound for any distribution whatsoever, and can therefore be used even when no information about the distribution except its statistics (mean and variance) are known.
2. Chebyshev's inequality is the best we can do when only the statistics of the distribution are known (by the previous example).
3. Chebyshev's inequality is still useful enough to provide very general and powerful theorems, such as the Weak Law of Large Numbers. These results do not depend on finer determination of the probability of the "tail" of the distribution.

Example

Example. Let X be a random variable with variance $\sigma^2 = 0$ and mean μ.

☞ By Chebyshev's inequality, the probability that X lies outside the interval $(\mu - \varepsilon, \mu + \varepsilon)$ equals zero for any $\varepsilon > 0$:

$$\mathbf{P}\left\{ |X - \mu| \geq \varepsilon \right\} \;\leq\; \frac{\sigma^2}{\varepsilon^2} \;=\; 0.$$

So, the event $\{X = \mu\}$ has probability 1.

☞ Let X be a random variable for which $E[X^2] = E[X]^2$. Then

$$Var(X) \;=\; E[X^2] - E[X]^2 \;=\; 0.$$

Thus, $\mathbf{P}\{X = E[X]\} = 1$.

Example

Example. Chebyshev's inequality provides a bound on the probability that a distribution lies greater than $k\sigma$ (k standard deviations) from its mean.

☞ Let X be any random variable with mean μ and variance σ^2.
Use $\varepsilon = k\sigma$ in Chebyshev:

$$\mathbf{P}\left\{|X - \mu| \geq k\sigma\right\} \;\leq\; \frac{\sigma^2}{k^2\sigma^2} \;=\; \frac{1}{k^2}.$$

☞ Here is a comparison of Chebyshev's bound with the known probability for a normally distributed random variable:

k	Chebyshev $(= \frac{1}{k^2})$	Normal
1	1	0.3164
2	0.5	0.0456
3	0.111	0.0038
4	0.0625	0.0000633

Example

Example. Let X be the number of heads obtained in tossing a fair coin 100 times. Find a bound on $\mathbf{P}\{41 \leq X \leq 59\}$.

The statistics for X are

$$E[X] = 100 \cdot \frac{1}{2} = 50 \qquad Var(X) = 100 \cdot \left(\frac{1}{2}\right)^2 = 25$$

☞ Use Chebyshev's inequality to obtain the bound:

$$\mathbf{P}\{41 \leq X \leq 59\} \; = \; 1 - \mathbf{P}\{|X - 50| \geq 10\} \; \geq \; 1 - \frac{25}{10^2} = 0.75.$$

☞ Using the standard normal approximation to the binomial distribution:

$$\begin{aligned} \mathbf{P}\{41 \leq X \leq 59\} &= \mathbf{P}\{40.5 \leq X \leq 59.5\} \\ &\approx \Phi\left(\frac{59.5 - 50}{5}\right) - \Phi\left(\frac{40.5 - 50}{5}\right) \approx 0.9426. \end{aligned}$$

☞ The actual probability is about 0.9431.

Problem

Problem. We wish to determine the value of an unknown X by making a measurement. However, we expect some variation in the expected measurement do to measuring inaccuracies.

Here are some examples

- X is the unknown bias of a coin.
- X is the angle of reflection in the collision of billiard balls.
- X is the average score on the final exam of a typical class in Probability theory.

Problem

☞ Let X have mean μ and variance σ^2. Chebyshev's inequality puts a bound on the accuracy ε of a single measurement of X

$$\mathbf{P}\{|X - \mu| < \varepsilon\} \; \geq \; 1 - \frac{\sigma^2}{\varepsilon^2}.$$

☞ If the variation σ is very small compared to the desired accuracy ε, then this probability will be very close to 1. That is, we can be “almost certain" that

$$X - \varepsilon \; \leq \; \mu \; \leq \; X + \varepsilon.$$

☞ If σ is not small compared to ε, then a single measurement cannot provide the desired accuracy with any certain.

We must either give-up accuracy on a single measurement ......

Problem

☞ ...... or else make a sufficiently large number of independent measurements.

☞ A powerful consequence of Chebyshev's inequality is that if we average of sufficiently many independent measurements of X, then the sample mean, $\overline{X}$, will with high probability provide the desired accuracy in the measurement:

$$\overline{X} - \varepsilon \; \leq \; \mu \; \leq \; \overline{X} + \varepsilon.$$

The reason is that $\overline{X}$ has the same mean μ, but smaller variance $\overline{\sigma^2}$.

Sampling reduces variation!!

☞ This result is the Weak Law of Large Numbers.

Definition IID random variables

Definition

A finite (or infinite) sequence of random variables $X_1, X_2, \ldots$ is said to be independent and identically distributed (IID) if they are mutually independent and each has the same distribution function:

$$\mathbf{P}\{X_i \le x\} = F(x) \qquad \text{for all } i \text{ and } x$$

Note that IID random variables have the same statistics: they have the same mean μ and variance σ^2:

$$E[X_i] = \mu \qquad Var(X_i) = \sigma^2 \qquad \text{for all } i$$

Mean and variance of Sample Mean

Proposition

Let $X_1, X_2, \ldots$ be IID random variables with finite common mean μ and variance σ^2.
Define the random variables: sum S_n and average A_n for each n by

$$S_n = X_1 + X_2 + \ldots + X_n \qquad A_n = \frac{S_n}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Then

$$E[S_n] = n \cdot \mu \qquad Var(S_n) = n \cdot \sigma^2$$
$$E[A_n] = \mu \qquad Var(A_n) = \frac{\sigma^2}{n}$$

Note: A_n is called the sample mean of $X_1, \ldots, X_n$ by statisticians and often denoted by $\overline{X}$.

Proof

☞ By the linearity property of expectation (no independence required):

$$E[S_n] = E[X_1] + \ldots + E[X_n] = n \cdot \mu$$
$$E[A_n] = E[\frac{S_n}{n}] = \frac{1}{n}E[S_n] = \mu.$$

☞ Since $X_1, \ldots, X_n$ are independent, variance is linear:

$$Var(S_n) = Var(X_1) + \ldots + Var(X_n) = n \cdot \sigma^2$$
$$Var(A_n) = Var(\frac{S_n}{n}) = \frac{1}{n^2} Var(S_n) = \frac{\sigma^2}{n}.$$

Example

Example. Let $X_1, X_2, \ldots, X_n$ be the indicator variables for the individual trials in a Bernoulli trials process with probability 0.3 for success.
So,

$$E[X_i] = 0.3 \qquad Var(X_i) = (0.3)(0.7) = 0.21.$$

☞ The sample average of the X_i is

$$A_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

So,

$$\mu = E[A_n] = 0.3 \qquad \sigma^2 = Var(A_n) = \frac{0.21}{n}.$$

Chebyshev's inequality for $\varepsilon = 0.1$ provides

$$\mathbf{P}\{0.2 \le A_n \le 0.4\} \ge 1 - \frac{0.21}{n(0.1)^2} = \frac{n-21}{n}.$$

Example – continued

$$\mathbf{P}\left\{0.2 \le A_n \le 0.4\right\} \ \ge\ 1 - \frac{21}{n} \ =\ \frac{n-21}{n}.$$

Bounds for $n = 50$, $n = 100$ and $n = 1000$:

$$\begin{aligned}\mathbf{P}\left\{0.2 \le A_{50} \le 0.4\right\} &\ge 0.58\\ \mathbf{P}\left\{0.2 \le A_{100} \le 0.4\right\} &\ge 0.79\\ \mathbf{P}\left\{0.2 \le A_{1000} \le 0.4\right\} &\ge 0.979.\end{aligned}$$

☞ The actual values are

$$\begin{aligned}\mathbf{P}\left\{0.2 \le A_{50} \le 0.4\right\} &\approx 0.836347\\ \mathbf{P}\left\{0.2 \le A_{100} \le 0.4\right\} &\approx 0.962549\\ \mathbf{P}\left\{0.2 \le A_{1000} \le 0.4\right\} &\approx 1.\end{aligned}$$

Example

Example. Suppose we choose at random n numbers in the interval $[0,1]$ with uniform distribution. Let X_i be the ith choice. Then

$$\begin{aligned}\mu &= E[X_i] = \int_0^1 x\,dx = \frac{1}{2}\\ \sigma^2 &= Var(X_i) = \int_0^1 x^2\,dx - \mu^2 = \frac{1}{12}.\end{aligned}$$

Let $A_n = \frac{X_1+\ldots+X_n}{n}$ be the sample average. Then

$$\begin{aligned}E\left[A_n\right] &= \frac{1}{2}\\ Var\left(A_n\right) &= \frac{1}{12n}\end{aligned}$$

By Chebyshev's Inequality, for any $\varepsilon > 0$

$$\mathbf{P}\left\{\left|A_n - \frac{1}{2}\right| < \varepsilon\right\} \ \ge\ 1 - \frac{1}{12n\varepsilon^2}.$$

Example – continued

$$\mathbf{P}\left\{\left|A_n - \frac{1}{2}\right| < \varepsilon\right\} \ \ge\ 1 - \frac{1}{12n\varepsilon^2}.$$

☞ This says that if we choose n numbers at random from $[0,1]$, then the chances are better than $1 - \frac{1}{12n\varepsilon^2}$ that the average of the chosen values differs from $\frac{1}{2}$ by less than ε.

☞ Suppose $\varepsilon = 0.1$. Then

$$\mathbf{P}\left\{\left|A_n - \frac{1}{2}\right| < \varepsilon\right\} \ \ge\ 1 - \frac{100}{12n}.$$

- For $n = 100$, the probability is about 0.92,
- For $n = 1000$, the probability is about 0.99,
- For $n = 10,000$, the probability is about 0.999.

Example

Example. Suppose we choose at random n numbers using the standard normal distribution. Let X_i be the ith choice. Then

$$\begin{aligned}\mu &= E[X_i] = 0\\ \sigma^2 &= Var(X_i) = 1.\end{aligned}$$

Let $A_n = \frac{X_1+\ldots+X_n}{n}$ be the sample average.
This is a normal distribution with mean 0 and variance 1. The statistics for the average value are

$$\begin{aligned}E\left[A_n\right] &= 0\\ Var\left(A_n\right) &= \frac{1}{n}\end{aligned}$$

By Chebyshev's Inequality, for any $\varepsilon > 0$

$$\mathbf{P}\left\{\left|A_n - 0\right| \ge \varepsilon\right\} \ \le\ \frac{1}{n\varepsilon^2}.$$

Example – continued

☞ For $\varepsilon = 0.1$ Chebyshev gives:

$$\mathbf{P}\left\{|A_n| \geq 0.1\right\} \leq \frac{100}{n}.$$

☞ Here is a comparison of the actual probabilities (A_n is normally distributed with mean 0 and variance n^{-1}) with the Chebyshev estimates:

n	$\mathbf{P}\left\{|A_n| \geq 0.1\right\}$	Chebyshev
100	0.3173	1.0000
200	0.1573	0.5000
300	0.0833	0.3333
400	0.0455	0.2500
500	0.0254	0.2000
600	0.0143	0.1667
700	0.0082	0.1429
800	0.0047	0.1250
900	0.0027	0.1111
1000	0.0016	0.1000

The Weak Law of Large Numbers

Theorem

Let $X_1, X_2, \ldots$ be a sequence of IID random variables with finite mean $E[X_i] = \mu$ and finite variance $Var(X_i) = \sigma^2$.

Let $A_n = \frac{X_1 + \ldots + X_n}{n}$ be the sample average. For any $\varepsilon > 0$,

$$\mathbf{P}\left\{ |A_n - \mu| \geq \varepsilon \right\} \to 0 \qquad as \qquad n \to \infty.$$

Equivalently,

$$\mathbf{P}\left\{ |A_n - \mu| < \varepsilon \right\} \to 1 \qquad as \qquad n \to \infty.$$

Proof

☞ Let $X_1, X_2, \ldots$ be IID. Then for the sample average $A_n = \frac{X_1 + \ldots + X_n}{n}$,

$$E[A_n] = \mu \qquad Var(A_n) = \frac{\sigma^2}{n}.$$

By Chebyshev's Inequality, for any fixed $\varepsilon > 0$,

$$\mathbf{P}\left\{ |A_n - \mu| \geq \varepsilon \right\} \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Thus, for fixed ε,

$$\mathbf{P}\left\{ |A_n - \mu| \geq \varepsilon \right\} \to 0 \qquad \text{as} \qquad n \to \infty.$$

Or equivalently,

$$\mathbf{P}\left\{ |A_n - \mu| < \varepsilon \right\} \to 1 \qquad \text{as} \qquad n \to \infty.$$

Measurements

☞ We want to make a measurement X with expected value μ and with some accuracy ϵ, where each measurement randomly varies with variance σ^2.

We also want a high degree of certainty of accuracy:

$$\mathbf{P}\left\{X - \epsilon \leq \mu \leq X + \varepsilon\right\} \geq p$$

☞ By Chebyshev's inequality, choose n sufficiently large so that

$$1 - p \geq \frac{\sigma^2}{n\varepsilon^2} \quad \text{equivalently} \quad n \geq \frac{\sigma^2}{(1-p)\varepsilon^2}.$$

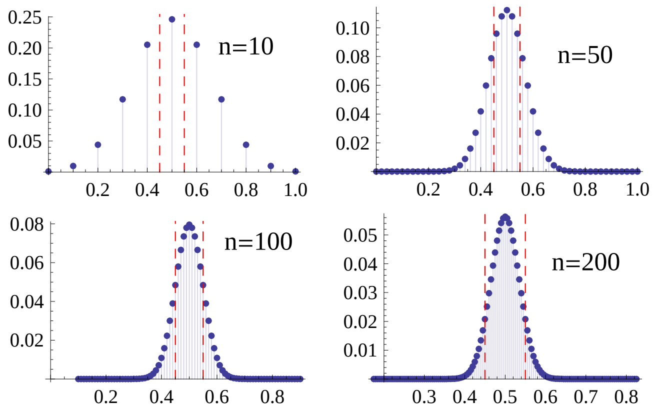
☞ To increase the accuracy $\varepsilon \to \frac{\varepsilon}{k}$, increase the trials by k^2:

$$n \geq \frac{\sigma^2}{(1-p)\varepsilon^2}. \quad \Longrightarrow \quad k^2 n \geq \frac{\sigma^2}{(1-p)\frac{\varepsilon}{k}^2}.$$

To Increase the accuracy by 1 decimal place, multiply trials by 100.

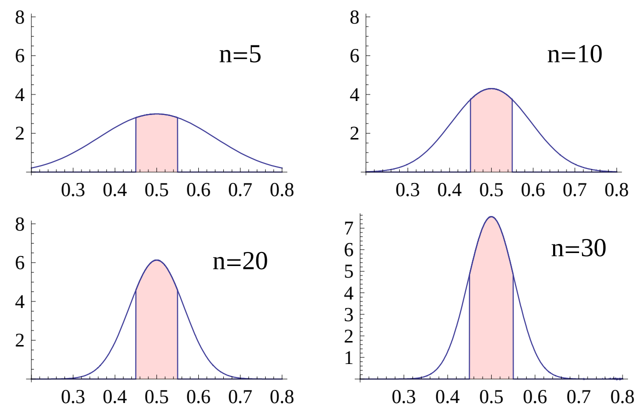
Example: Coin Tossing

☞ Plot of the mass of the random variable A_n (sample average) of a fair coin tossed n times. The mean of A_n is $\mu = 0.5$. Note the larger n, the greater the percentage of area is contained in the interval $(0.45, 0.55)$, as predicted by the Law of Large Numbers.



Example: Random Numbers

☞ Plot of the density of the random variable A_n (sample average) of n reals uniformly chosen in $[0, 1]$. The mean of A_n is $\mu = 0.5$. Note the larger n, the greater the percentage of area is contained in the interval $(0.45, 0.55)$, as predicted by the Law of Large Numbers.



Example – continued

☞ Consider a Bernoulli trials process with IID indicator variables $X_1, X_2, \ldots$ denoting whether the trial was a success or failure. Suppose the probability of success is p. So,

$$E[X_i] = p \qquad Var(X_i) = p(1-p).$$

☞ Let $A_n = \frac{X_1 + \ldots + X_n}{n}$ be the same average over n trials. So,

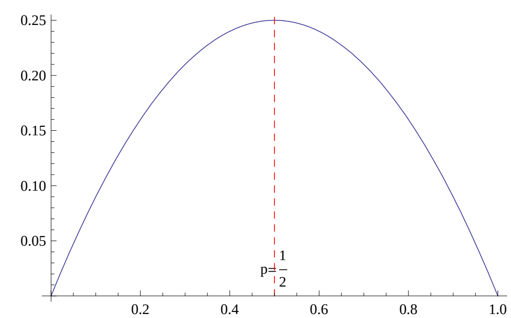
$$E\left[A_n\right] = p \qquad Var\left(A_n\right) = \frac{p(1-p)}{n}.$$

☞ By Chebyshev's inequality, for any $\varepsilon > 0$

$$\mathbf{P}\left\{\left|A_n - p\right| \geq \varepsilon\right\} \leq \frac{p(1-p)}{n\varepsilon^2}.$$

Example – continued

Example – continued. The variance $\sigma^2 = p(1-p)$ has a maximum value of $\frac{1}{4}$ achieved at $p = \frac{1}{2}$:



Plugging back into Chebyshev gives us a bound on the deviation of the sample average from the mean:

$$\mathbf{P}\left\{\left|A_n - p\right| \geq \varepsilon\right\} \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

Example

Example

We have two coins: one is fair and the other produces heads with probability $3/4$. One coin is picked at random. How many tosses suffice for us to be 95 percent sure of which coin we had?

☞ To make this problem more concrete: if the proportion of heads is less than 0.625, then we will guess the coin was fair; otherwise, if the proportion of heads is greater than 0.625 we will guess the biased coin.
How many tosses will suffice for 95 percent certainty that the chosen coin will not deviate by more than $\epsilon = 0.125$ from its mean?

Example – continued

☞ We have the following bound on the deviation of the sample average A_n from the mean p using Chebyshev's inequality:

$$\mathbf{P}\left\{|A_n - p| \geq \varepsilon\right\} \leq \frac{1}{4n\varepsilon^2}.$$

☞ We want to n large enough so that we have only 5% error:

$$\frac{1}{4n\varepsilon^2} \leq 0.05$$

equivalently,

$$n \geq \frac{1}{4(0.05)\varepsilon^2}.$$

☞ We now have a bound on the number of trials needed without needing to know the mean or the variance.

Example – continued

☞ For $\varepsilon = 0.125$ choose n so that

$$n \geq \frac{1}{4(0.05)(0.125)^2} \quad \text{equivalently} \quad n \geq 320.$$

☞ By tossing the coin $n \geq 320$ times we can be 95% certain the sample average is within 0.125 of the true bias p of the coin to heads:

$$\mathbf{P}\left\{|A_n - p| \geq 0.125\right\} \leq 0.05$$

☞ Toss the coin 320 times and count heads.

- If fewer than 200 heads appear guess the fair coin.
- If more than 200 heads appear guess the biases coin.
- If exactly 200 heads appears, then laugh at your (bad?) luck.

You can be 95 percent certain you chose the right coin.

Degree of Certainty vs. Number of Trials

☞ To achieve certainty p that we are within $\varepsilon = 0.125$ of the mean requires n trials, where

$$n \geq \frac{1}{4(1-p)(0.125)^2}$$

Degree of Certainty	Number of Trials
50%	32
75%	64
90%	160
95%	320
99%	1600
99.9%	16,000