Discrete Mathematics & Mathematical Reasoning
Chapter 7 (continued): Markov and Chebyshev's Inequalities; and Examples in probability: the birthday problem

Kousha Etessami

U. of Edinburgh, UK

Markov's Inequality

Often, for a random variable X that we are interested in, we want to know

> *"What is the probability that the value of the r.v., X, is 'far' from its expectation?"*

A generic answer to this, which holds for any non-negative random variable, is given by Markov's inequality:

Markov's Inequality

Theorem: For a nonnegative random variable, $X : \Omega \rightarrow \mathbb{R}$, where $X(s) \geq 0$ for all $s \in \Omega$, for any positive real number $a > 0$:

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Proof of Markov's Inequality:

Let the event $A \subseteq \Omega$ be defined by: $A = \{s \in \Omega \mid X(s) \geq a\}$.

We want to prove that $P(A) \leq \frac{E(X)}{a}$. But:

$$\begin{aligned}
E(X) &= \sum_{s \in \Omega} P(s)X(s) \\
&= \sum_{s \in A} P(s)X(s) + \sum_{s \notin A} P(s)X(s) \\
&\geq \sum_{s \in A} P(s)X(s) \qquad \text{(because } X(s) \geq 0 \text{ for all } s \in \Omega\text{)} \\
&\geq \sum_{s \in A} P(s)a \qquad \text{(because } X(s) \geq a \text{ for all } s \in A\text{)} \\
&= a \sum_{s \in A} P(s) \;=\; a \cdot P(A)
\end{aligned}$$

Thus, $E(X) \geq a \cdot P(A)$. In other words, $\frac{E(X)}{a} \geq P(A)$, which is what we wanted to prove.

Example

Question: A biased coin, which lands heads with probability $1/10$ each time it is flipped, is flipped 200 times consecutively. Give an upper bound on the probability that it lands heads at least 120 times.

Example

Question: A biased coin, which lands heads with probability $1/10$ each time it is flipped, is flipped 200 times consecutively. Give an upper bound on the probability that it lands heads at least 120 times.

Answer: The number of heads is a binomially distributed r.v., X, with parameters $p = 1/10$ and $n = 200$.

Thus, the expected number of heads is
$E(X) = np = 200 \cdot (1/10) = 20$.

By Markov Inequality, the probability of at least 120 heads is

$$P(X \geq 120) \leq \frac{E(X)}{120} = \frac{20}{120} = 1/6. \quad \square$$

Later we will see that one can give MUCH MUCH BETTER bounds in this specific case.

Chebyshev's Inequality

Another answer to the question of "what is the probability that the value of X is far from its expectation" is given by Chebyshev's Inequality, which works for any random variable (not necessarily a non-negative one).

Chebyshev's Inequality

Theorem: Let $X : \Omega \rightarrow \mathbb{R}$ be any random variable, and let $r > 0$ be any positive real number. Then:

$$P(|X - E(X)| \geq r) \leq \frac{V(X)}{r^2}$$

First proof of Chebyshev's Inequality:

Let $A \subseteq \Omega$ be defined by: $A = \{s \in \Omega \mid |X(s) - E(X)| \geq r\}$.

We want to prove that $P(A) \leq \frac{V(X)}{r^2}$. But:

$$\begin{aligned}
V(X) &= \sum_{s \in \Omega} P(s)(X(s) - E(X))^2 \\
&= \sum_{s \in A} P(s)(X(s) - E(X))^2 + \sum_{s \notin A} P(s)(X(s) - E(X))^2 \\
&\geq \sum_{s \in A} P(s)(X(s) - E(X))^2 \text{ (since } \forall s, (X(s) - E(X))^2 \geq 0) \\
&\geq \sum_{s \in A} P(s) r^2 \text{ (because } |X(s) - E(X)| \geq r \text{ for all } s \in A) \\
&= r^2 \sum_{s \in A} P(s) \; = \; r^2 \cdot P(A)
\end{aligned}$$

Thus, $V(X) \geq r^2 \cdot P(A)$. In other words, $\frac{V(X)}{r^2} \geq P(A)$, which is what we wanted to prove.

Our first proof of Chebyshev's inequality looked suspiciously like our proof of Markov's Inequality. That is no co-incidence. Chebyshev's inequality can be derived as a special case of Markov's inequality.

Second proof of Chebyshev's Inequality:

Note that
$A = \{s \in \Omega \mid |X(s) - E(X)| \geq r\} = \{s \in \Omega \mid (X(s) - E(X))^2 \geq r^2\}$.
Now, consider the random variable, Y, where
$Y(s) = (X(s) - E(X))^2$.
Note that Y is a non-negative random variable.
Thus, we can apply Markov's inequality to it, to get:

$$P(A) = P(Y \geq r^2) \leq \frac{E(Y)}{r^2} = \frac{E((X - E(X))^2)}{r^2} = \frac{V(X)}{r^2}. \quad \square$$

Brief look at a more advanced topic: Chernoff bounds

For specific random variables, particularly those that arise as *sums of many independent random variables*, we can get much better bounds on the probability of deviation from expectation.

Some special cases of **Chernoff Bounds**:

Theorem: Suppose we conduct a sequence of n mutually independent Bernoulli trials, $X_i \in \{0, 1\}$, with probability p of "success" (i.e., getting a 1, i.e., heads) in each trial. Let $X = \sum_{i=1}^{n} X_i$ be the binomially distributed r.v. that counts the total number of successes (recall that $E(X) = pn$). Then:

1. For all $\epsilon > 0$, $P(|X - E(X)| \geq \epsilon n) \leq 2e^{-2n\epsilon^2}$.
2. For all $C \geq 6E(X)$, $P(X \geq C) \leq 2^{-C}$.

We will not prove this theorem, and we will not assume you know it (it is not in the book).

An application of Chernoff bounds

Question: A biased coin is flipped 200 times consecutively, and comes up heads with probability $1/10$ each time it is flipped. Give an upper bound the probability that it will come up heads at least 120 times.

An application of Chernoff bounds

Question: A biased coin is flipped 200 times consecutively, and comes up heads with probability $1/10$ each time it is flipped. Give an upper bound the probability that it will come up heads at least 120 times.

Solution: Let X be the r.v. that counts the number of heads. Recall: $E(X) = 200 * (1/10) = 20$. By Chernoff bounds,

$$P(X \geq 120) = P(X \geq 6E(X)) \leq 2^{-6E(X)} = 2^{-(6 \cdot 20)} = 2^{-120}. \quad \square$$

Note: By using Markov's inequality, we were only able to determine that $P(X \geq 120) \leq (1/6)$.
But by using Chernoff bounds, which are specifically geared for large deviation bounds for binomial and related distributions, we get that $P(X \geq 120) \leq 2^{-120}$.
That is a vastly better upper bound!

Another application of Chernoff bounds: "how many random samples do I need?"

Suppose you are given a biased coin, which lands heads with probability p each time it is flipped. **But** you are not told what p is. You want to find out what p is.
To estimate p, you can of course flip the coin n times, count the number of times, X, that it lands heads, and give the estimate:

$$\text{"}p \text{ is roughly } \frac{X}{n}.\text{"}$$

But how big does n have to be for your estimate $\frac{X}{n}$ to (probably) be "good"?

Another application of Chernoff bounds: “how many random samples do I need?”

Suppose you are given a biased coin, which lands heads with probability p each time it is flipped. **But** you are not told what p is. You want to find out what p is.

To estimate p, you can of course flip the coin n times, count the number of times, X, that it lands heads, and give the estimate:

$$\text{“}p \text{ is roughly } \frac{X}{n}.\text{”}$$

But how big does n have to be for your estimate $\frac{X}{n}$ to (probably) be “good”? Concretely, how many independent random samples (coin flips), n, do you need in order to be sure that, say:

$$P(|\frac{X}{n} - p| > \frac{1}{30}) \leq \frac{1}{25} \quad ?$$

Note that this is equivalent to: $P(|X - pn| > \frac{1}{30}n) \leq \frac{1}{25}$.

We can use Chernoff bounds to bound n.

Question: How many random samples n do I need to make sure that:

$$P(|X - pn| > \frac{1}{30}n) \leq \frac{1}{25} \quad ?$$

Solution: Let X be the r.v. that counts the number of heads from n coin flips. Recall that $E(X) = pn$.
By Chernoff bounds, $P(|X - pn| \geq \epsilon n) \leq 2e^{-2n\epsilon^2}$.
Let $\epsilon = \frac{1}{30}$, and let $n = 1800$, then

$$P(|X - pn| \geq \tfrac{1}{30}n) \leq 2e^{-2\cdot 1800\cdot(\frac{1}{30})^2} = 2e^{-4} = 0.0366 \leq \tfrac{1}{25}.$$

Thus, 1800 random samples (i.e., coin flips) suffice to make sure that, with probability at least $\frac{24}{25}$, your estimate $\frac{X}{n}$ for the coin's bias p is correct to within additive error at most $\frac{1}{30}$.

These kinds of bounds are crucial in statistical analysis.

The Birthday Problem

There are many illuminating and surprising examples in probability theory. We will see some of them in the next couple of lectures, in order to build our intuition about probability.

One well-known example is called the Birthday problem.

Birthday problem

There are 30 people in a room. I am willing to bet you that "at least two people in the room have the same birthday".

Should you take my bet? (I offer even odds.)

The Birthday Problem

There are many illuminating and surprising examples in probability theory. We will see some of them in the next couple of lectures, in order to build our intuition about probability.

One well-known example is called the Birthday problem.

Birthday problem

There are 30 people in a room. I am willing to bet you that "at least two people in the room have the same birthday".

Should you take my bet? (I offer even odds.)

In order words, you have to calculate:
is there at least $1/2$ probability that no two people will have the same birthday in a room with 30 people?
(We are implicitly assuming that these people's birthdays are independent and uniformly distributed throughout the 365(+1) days of the year, taking into account leap years.)

Toward a solution to the Birthday problem:

Question: What is the probability, p_m, that m people in a room all have different birthdays?

Toward a solution to the Birthday problem:

Question: What is the probability, p_m, that m people in a room all have different birthdays?

We can equate the birthdays of m people to a list $(b_1, \ldots, b_m)$, with each $b_i \in \{1, \ldots, 366\}$.

We are assuming each list in $B = \{1, \ldots, 366\}^m$ is equally likely.

Note that $|B| = 366^m$. What is the size of

$A = \{(b_1, \ldots, b_m) \in B \mid b_i \neq b_j \text{ for all } i \neq j, i, j \in \{1, \ldots, m\}\}$?

This is simply the # of m-permutations from a set of size 366.

Thus $|A| = 366 \cdot (366 - 1) \ldots (366 - (m - 1))$.

Thus, $p_m = \frac{|A|}{|B|} = \prod_{i=1}^{m} \frac{366-i+1}{366} = \prod_{i=1}^{m} (1 - \frac{i-1}{366})$.

By brute-force calculation, $p_{30} = 0.2947$. Thus, the probability that at least two people do have the same birthday in a room with 30 people is $1 - p_{30} = 0.7053$.

So, you shouldn't have taken my bet! Not even for 23 people in a room, because $1 - p_{23} = 0.5063$. But $1 - p_{22} = 0.4745$.

A general result underlying the birthday paradox

Theorem: Suppose that each of $m \geq 1$ pigeons independently and uniformly at random enter one of $n \geq 1$ pigeon-holes. If

$$m \geq (1.1775 \cdot \sqrt{n}) + 1$$

then the probability that two pigeons go into the same pigeon-hole is greater than $1/2$.

Proof: **Basic Fact:** $1 + x \leq e^x$, for all real numbers x.
The probability that m random pigeons all go in different pigeonholes, when there are n pigeonholes, is:

$$\prod_{i=1}^{m-1}(1 + (-\frac{i}{n})) \leq \prod_{i=1}^{m-1} e^{-(i/n)} = e^{-\frac{1}{n}\sum_{i=1}^{m-1} i} = e^{-\frac{m(m-1)}{2n}}$$

So we want m to be big enough so that $e^{-\frac{m(m-1)}{2n}} < 1/2$.
Taking logs, and negating, this is equivalent to

$$\frac{m(m-1)}{2n} > \ln 2 \iff m(m-1) > (2 \cdot \ln 2) \cdot n$$

Thus, since $m(m-1) > (m-1)^2$, it suffices if

$$(m-1)^2 \geq (2 \cdot \ln 2) \cdot n \iff (m-1) \geq \sqrt{(2 \cdot \ln 2)} \cdot \sqrt{n}$$

Thus, since $\sqrt{(2 \ln 2)} = 1.177410 \ldots \leq 1.1775$, it suffices if:

$$m \geq (1.1775 \cdot \sqrt{n}) + 1. \quad \square$$

We will not assume you know this proof.