CMPT 210: Probability and Computing

Lecture 21

Sharan Vaswani

March 24, 2023

Recap

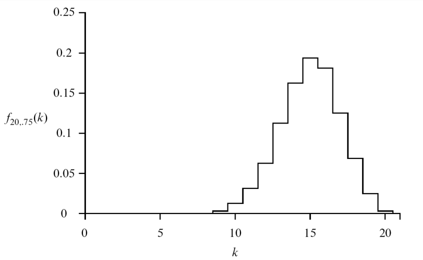
Tail inequalities bound the probability that the r.v. takes a value much different from its mean.

Markov's Theorem: If X is a non-negative random variable, then for all $x > 0$, $\Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}$.

Chebyshev's Theorem: For a r.v. X and all $x > 0$, $\Pr[|X - \mathbb{E}[X]| \geq x] \leq \frac{\mathrm{Var}[X]}{x^2}$.

Chebyshev's Theorem - Example

Q: Consider a r.v. $X \sim \text{Bin}(20, 0.75)$. Plot the PDF$_X$, compute its mean and standard deviation and bound $\Pr[10 < X < 20]$.



Range$(X) = \{0, 1, \ldots, 20\}$ and for $k \in$ Range(X),
$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.
$\mathbb{E}[X] = np = (20)(0.75) = 15$
$\text{Var}[X] = np(1-p) = 20(0.75)(0.25) = 3.75$ and hence
$\sigma_X = \sqrt{3.75} \approx 1.94$.

$$\begin{aligned}
\Pr[10 < X < 20] &= 1 - \Pr[X \leq 10 \cup X \geq 20] \\
&= 1 - \Pr[|X - 15| \geq 5] \\
&= 1 - \Pr[|X - \mathbb{E}[X]| \geq 5] \\
&\geq 1 - \frac{\text{Var}[X]}{(5)^2} = 1 - \frac{3.75}{25} = 0.85.
\end{aligned}$$

Hence, the "probability mass" of X is "concentrated" around its mean.

Voter Poll

Q: Suppose there is an election between two candidates Donald Trump and Joe Biden, and we are hired by candidate Biden's election campaign to estimate his chances of winning the election. In particular, we want to estimate p, the fraction of voters favoring Biden before the election. We conduct a voter poll – selecting (typically calling) people uniformly at random (with replacement so that we can choose a person twice) and try to estimate p. What is the number of people we should poll to estimate p reasonably accurately and with reasonably high probability?

Define X_i to be the indicator r.v. equal to 1 iff person i that we called favors Biden.

Assumption (1): The X_i r.v's are mutually independent since the people we poll are chosen randomly and we assume that their opinions do not affect each other.

Assumption (2): The people we call are identically distributed i.e. $X_i = 1$ with probability p.

Suppose we poll n people and define $S_n := \sum_{i=1}^{n} X_i$ as the r.v. equal to the total number of people (amongst the ones we polled) that prefer Biden. $\frac{S_n}{n}$ is the *statistical estimate* of p.

Q: What is the distribution of S_n?

Voter Poll

We want to find for what n is our estimate for p accurate up to an error $\epsilon > 0$ and with probability $1 - \delta$ (for $\delta \in (0, 1)$). Formally, for what n is,

$$\Pr\left[\left|\frac{S_n}{n} - p\right| < \epsilon\right] \geq 1 - \delta$$

Since $S_n \sim \text{Bin}(n, p)$, $\mathbb{E}[S_n] = np$ and hence, $\mathbb{E}\left[\frac{S_n}{n}\right] = p$, meaning that our estimate is *unbiased* – in expectation, the estimate is equal to p. Hence, the above statement is equivalent to,

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| < \epsilon\right] \geq 1 - \delta$$

Hence, we can use Chebyshev's Theorem for the r.v. $\frac{S_n}{n}$ with $x = \epsilon$ to bound the LHS

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| < \epsilon\right] = 1 - \Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| \geq \epsilon\right] \geq 1 - \frac{\text{Var}[S_n/n]}{\epsilon^2}.$$

Hence, the problem now is to find n such that,

$$1 - \frac{\text{Var}[S_n/n]}{\epsilon^2} \geq 1 - \delta \implies \frac{\text{Var}[S_n/n]}{\epsilon^2} < \delta$$

Voter Poll

Let us calculate the $\mathrm{Var}[S_n/n]$.

$$\mathrm{Var}[S_n/n] = \frac{1}{n^2}\mathrm{Var}[S_n] \qquad \text{(Using the property of variance)}$$

$$= \frac{1}{n^2}\, n\, p\,(1-p) = \frac{p\,(1-p)}{n} \qquad \text{(Using the variance of the Binomial distribution)}$$

Hence, we want to find n s.t.

$$\frac{p\,(1-p)}{n\epsilon^2} < \delta \implies n \geq \frac{p(1-p)}{\epsilon^2\,\delta}$$

But we do not know p! If $n \geq \max_p \frac{p(1-p)}{\epsilon^2\,\delta}$, then for any p, $n \geq \frac{p(1-p)}{\epsilon^2\,\delta}$. So the problem is to compute $\max_p \frac{p(1-p)}{\epsilon^2\,\delta}$. This is a concave function and is maximized at $p = 1/2$.

Hence, if $n \geq \frac{1}{4\epsilon^2\delta}$, then $\Pr\left[\left|\frac{S_n}{n} - p\right| < \epsilon\right] \geq 1 - \delta$ meaning that we have estimated p upto an error ϵ and this bound is true with high probability equal to $1 - \delta$.

For example, if $\epsilon = 0.01$ and $\delta = 0.01$ meaning that we want the bound to hold 99% of the time, then, we require $n \geq 250000$.

Pairwise Independent Sampling

Claim: Let $G_1, G_2, \ldots, G_n$ be pairwise independent random variables with the same mean μ and standard deviation σ. Define $S_n := \sum_{i=1}^{n} G_i$, then,

$$\Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right] \leq \frac{1}{n}\left(\frac{\sigma}{\epsilon}\right)^2.$$

Proof: Let us compute $\mathbb{E}[S_n/n]$ and $\mathrm{Var}[S_n/n]$.

$$\mathbb{E}[S_n] = \mathbb{E}\left[\sum_{i=1}^{n} G_i\right] = \sum_{i=1}^{n} \mathbb{E}[G_i] = n\mu \implies \mathbb{E}[S_n/n] = \frac{1}{n}\mathbb{E}[S_n] = \mu$$

(Using linearity of expectation)

$$\mathrm{Var}[S_n] = \mathrm{Var}\left[\sum_{i=1}^{n} G_i\right] = \sum_{i=1}^{n} \mathrm{Var}[G_i] = n\sigma^2$$

(Using linearity of variance for pairwise independent r.v's)

$$\implies \mathrm{Var}[S_n/n] = \frac{1}{n^2}\mathrm{Var}[S_n] = \frac{\sigma^2}{n}$$

Pairwise Independent Sampling

Using Chebyshev's Theorem,

$$\Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right| \geq \epsilon\right] = \Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right] \leq \frac{\mathrm{Var}[S_n/n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

Hence, for arbitrary pairwise independent r.v's, if n increases, the probability of deviation from the mean μ decreases.

Weak Law of Large Numbers: Let $G_1, G_2, \ldots, G_n$ be pairwise independent variables with the same mean μ and (finite) standard deviation σ. Define $X_n := \frac{\sum_{i=1}^{n} G_i}{n}$, then for every $\epsilon > 0$,

$$\lim_{n \to \infty} \Pr[|X_n - \mu| \leq \epsilon] = 1.$$

Proof: Follows from the theorem on pairwise independent sampling since $\lim_{n\to\infty} \Pr[|X_n - \mu| \leq \epsilon] = \lim_{n\to\infty} \left[1 - \frac{\sigma^2}{n\epsilon^2}\right] = 1$.

Questions?

Sums of Random Variables

If we know that the r.v X is (i) non-negative and (ii) $\mathbb{E}[X]$, we can use Markov's Theorem to bound the probability of deviation from the mean.

If we know both (i) $\mathbb{E}[X]$ and (ii) $\mathrm{Var}[X]$, we can use Chebyshev's Theorem to bound the probability of deviation.

In many cases (the voter poll example), the random variable of interest is a sum of r.v's (e.g., for the voter poll application), and we can use the Chernoff bound to obtain tighter bounds on the deviation from the mean.

Chernoff Bound: Let $T_1, T_2, \ldots, T_n$ be mutually independent r.v's such that $0 \leq T_i \leq 1$ for all i. If $T := \sum_{i=1}^{n} T_i$, for all $c \geq 1$ and $\beta(c) := c\ln(c) - c + 1$,

$$\Pr[T \geq c\mathbb{E}[T]] \leq \exp(-\beta(c)\,\mathbb{E}[T])$$

If $T_i \sim \mathrm{Ber}(p)$ and are mutually independent, then $T_i \in \{0, 1\}$ and we can use the Chernoff bound to bound the deviation from the mean for $T \sim \mathrm{Bin}(n, p)$. In general, if $T_i \in [0, 1]$, the Chernoff Bound can be used even if the T_i's have different distributions!

Chernoff Bound – Binomial Distribution

Q: Bound the probability that the number of heads that come up in 1000 independent tosses of a fair coin exceeds the expectation by 20% or more.

Let T_i be the r.v. for the event that coin i comes up heads, and let T denote the total number of heads. Hence, $T = \sum_{i=1}^{1000} T_i$. For all i, $T_i \in \{0, 1\}$ and are mutually independent r.v's. Hence, we can use the Chernoff Bound.

We want to compute the probability that the number of heads is larger than the expectation by 20% meaning that $c = 1.2$ for the Chernoff Bound. Computing $\beta(c) = c\ln(c) - c + 1 \approx 0.0187$. Since the coin is fair, $\mathbb{E}[T] = 1000\,\frac{1}{2} = 500$. Plugging into the Chernoff Bound,

$$\Pr[T \geq c\mathbb{E}[T]] \leq \exp(-\beta(c)\mathbb{E}[T]) \implies \Pr[T \geq 1.2\,\mathbb{E}[T]] \leq \exp(-(0.0187)\,(500)) \approx 0.0000834.$$

Comparing this to using Chebyshev's inequality,

$$\begin{aligned}\Pr[T \geq c\mathbb{E}[T]] &= \Pr[T - \mathbb{E}[T] \geq (c-1)\mathbb{E}[T]] \leq \Pr[|T - \mathbb{E}[T]| \geq (c-1)\,\mathbb{E}[T]] \\ &\leq \frac{\mathrm{Var}[T]}{(c-1)^2\,(\mathbb{E}[T])^2} = \frac{1000\,\frac{1}{4}}{(1.2-1)^2(500^2)} = \frac{250}{0.2^2\,500^2} = \frac{250}{10000} = 0.025.\end{aligned}$$

Chernoff Bound – Lottery Game

Q: Pick-4 is a lottery game in which you pay \$1 to pick a 4-digit number between 0000 and 9999. If your number comes up in a random drawing, then you win \$5,000. Your chance of winning is 1 in 10000. If 10 million people play, then the expected number of winners is 1000. When there are 1000 winners, the lottery keeps \$5 million of the \$10 million paid for tickets. The lottery operator's nightmare is that the number of winners is much greater – especially at the point where more than 2000 win and the lottery must pay out more than it received. What is the probability that will happen? (Assume that the players' picks and the winning number are random, independent and uniform)

Let T_i be an indicator for the event that player i wins. Then $T := \sum_{i=1}^{n} T_i$ is the total number of winners. Using the independence assumptions, we canc conclude that T_i are independent, as required by the Chernoff bound.

We wish to compute $\Pr[T \geq 2000] = \Pr[T \geq 2\mathbb{E}[T]]$. Hence $c = 2$ and $\beta(c) \approx 0.386$. By the Chernoff bound,

$$\Pr[T \geq 2\mathbb{E}[T]] \leq \exp(-\beta(c)\mathbb{E}[T]) = \exp(-(0.386)\,1000) < \exp(-386) \approx 10^{-168}$$

Questions?

Chernoff Bound – Proof

Chernoff Bound: Let $T_1, T_2, \ldots, T_n$ be mutually independent r.v's such that $0 \le T_i \le 1$ for all i. If $T := \sum_{i=1}^{n} T_i$, for all $c \ge 1$ and $\beta(c) := c\ln(c) - c + 1$,

$$\Pr[T \ge c\mathbb{E}[T]] \le \exp(-\beta(c)\,\mathbb{E}[T])$$

Proof: We want to compute $\Pr[T \ge c\mathbb{E}[T]] = \Pr[f(T) \ge f(c\mathbb{E}[T])]$ where f is a one-one monotonically non-decreasing function. For $c \ge 1$, choosing $f(T) = c^T$ and using Markov's Theorem,

$$\begin{aligned}
\Pr[T \ge c\mathbb{E}[T]] = \Pr[c^T \ge c^{c\mathbb{E}[T]}] &\le \frac{\mathbb{E}[c^T]}{c^{c\mathbb{E}[T]}} \\
&\le \frac{\exp((c-1)\,\mathbb{E}[T])}{c^{c\mathbb{E}[T]}} \qquad \text{(To prove next: } \mathbb{E}[c^T] \le \exp((c-1)\,\mathbb{E}[T])\text{)} \\
&= \frac{\exp((c-1)\,\mathbb{E}[T])}{\exp(\ln(c^{c\mathbb{E}[T]}))} = \frac{\exp((c-1)\,\mathbb{E}[T])}{\exp(c\mathbb{E}[T]\,\ln(c))} = \exp\left(-(c\ln(c) - c + 1)\,\mathbb{E}[T]\right)
\end{aligned}$$

$$\implies \Pr[T \ge c\mathbb{E}[T]] \le \exp\left(-\beta(c)\,\mathbb{E}[T]\right)$$

The proof would be done if we prove that $\mathbb{E}[c^T] \le \exp((c-1)\,\mathbb{E}[T])$ and we do this next.

Chernoff Bound – Proof

Claim: $\mathbb{E}[c^T] \leq \exp((c-1)\,\mathbb{E}[T])$

$$\mathbb{E}[c^T] = \mathbb{E}[c^{\sum_{i=1}^n T_i}] = \mathbb{E}\left[\prod_{i=1}^n c^{T_i}\right] = \prod_{i=1}^n \mathbb{E}[c^{T_i}]$$

(Expectation of product of mutually independent r.v's is equal to the product of the expectation.)

$$\leq \prod_{i=1}^n \exp((c-1)\,\mathbb{E}[T_i]) \qquad \text{(To prove next: } \mathbb{E}[c^{T_i}] \leq \exp((c-1)\,\mathbb{E}[T_i]))$$

$$= \exp\left((c-1)\sum_{i=1}^n \mathbb{E}[T_i]\right) = \exp\left((c-1)\,\mathbb{E}\left[\sum_{i=1}^n T_i\right]\right)$$

(Linearity of Expectation)

$$\implies \mathbb{E}[c^T] \leq \exp((c-1)\mathbb{E}[T])$$

The proof would be done if we prove that $\mathbb{E}[c^{T_i}] \leq \exp((c-1)\,\mathbb{E}[T_i])$ and we do this next.

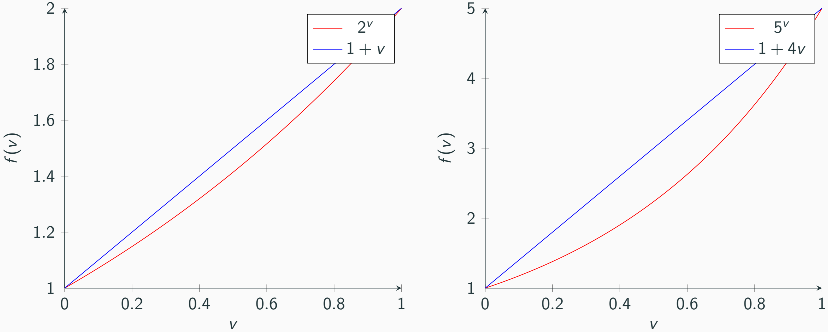
Chernoff Bound – Proof

Claim: $\mathbb{E}[c^{T_i}] \leq \exp((c-1)\,\mathbb{E}[T_i])$

$$\mathbb{E}[c^{T_i}] = \sum_{v \in \text{Range}(T_i)} c^v \Pr[T_i = v] \leq \sum_{v \in \text{Range}(T_i)} (1 + (c-1)v) \Pr[T_i = v]$$

(Since $T_i \in [0,1]$ and $c^v \leq 1 + (c-1)v$ for all $v \in [0,1]$.)

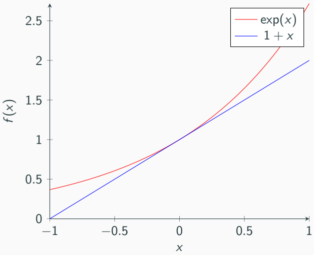
For $c = 2$ and $c = 5$,



Chernoff Bound – Proof

$$\begin{aligned}
\mathbb{E}[c^{T_i}] &\le \sum_{v \in \mathrm{Range}(T_i)} (1 + (c-1)v) \Pr[T_i = v] \\
&= \sum_{v \in \mathrm{Range}(T_i)} \Pr[T_i = v] + (c-1) \sum_{v \in \mathrm{Range}(T_i)} v \Pr[T_i = v] \\
&= 1 + (c-1)\,\mathbb{E}[T_i] \le \exp((c-1)\mathbb{E}[T_i]) \qquad \text{(Since } 1 + x \le \exp(x) \text{ for all } x\text{)}
\end{aligned}$$

$$\implies \mathbb{E}[c^{T_i}] \le \exp((c-1)\,\mathbb{E}[T_i])$$



Hence we have proved the Chernoff Bound!

Comparing the Bounds

For r.v's $T_1, T_2, \ldots T_n$, if $T_i \in \{0,1\}$ and $\Pr[T_i = 1] = p_i$. Define $T := \sum_{i=1}^n T_i$. By linearity of expectation, $\mathbb{E}[T] = \sum_{i=1}^n p_i$. For $c \geq 1$,

Markov's Theorem: $\Pr[T \geq c\mathbb{E}[T]] \leq \frac{1}{c}$. Does not require T_i's to be independent.

Chebyshev's Theorem:

$$\Pr[T - \mathbb{E}[T] \geq x] \leq \Pr[|T - \mathbb{E}[T]| \geq x] \leq \frac{\mathrm{Var}[T]}{x^2}$$

$$\implies \Pr[T - \mathbb{E}[T] \geq (c-1)\mathbb{E}[T]] \leq \frac{\mathrm{Var}[T]}{(c-1)^2\,(\mathbb{E}[T])^2)} \qquad (x = (c-1)\,\mathbb{E}[T])$$

If the T_i's are pairwise independent, by linearity of variance, $\mathrm{Var}[T] = \sum_{i=1}^n p_i\,(1-p_i)$. Hence, $\Pr[T \geq c\mathbb{E}[T]] \leq \frac{\sum_{i=1}^n p_i\,(1-p_i)}{(c-1)^2\left(\sum_{i=1}^n p_i\right)^2}$. If for all i, $p_i = 1/2$, then, $\Pr[T \geq c\mathbb{E}[T]] \leq \frac{1}{(c-1)^2}$.

Chernoff Bound: If T_i' are mutually independent, then, $\Pr[T \geq c\mathbb{E}[T]] \leq \exp(-\beta(c)\,\mathbb{E}[T]) = \exp\left(-(c\ln(c) - c + 1)\,\left(\sum_{i=1}^n p_i\right)\right)$. If for all i, $p_i = 1/2$, $\Pr[T \geq c\mathbb{E}[T]] \leq \exp\left(-\frac{n\,(c\ln(c)-c+1)}{2}\right)$.

Questions?