## Warm-up

**Problem 1.** Suppose $E_1$ and $E_2$ are two *independent* events, each happening with probability $p$. What is the probability that at least one of them happens? Compare to what the union bound gives.

Generalise to $k$ independent events $E_1, \ldots, E_k$ each happening with probability $p$.

**Solution 1.** The complementary events $\bar{E}_1$ and $\bar{E}_2$ are then independent as well, and easier to work with:

$$\Pr[\, E_1 \cup E_2 \,] = 1 - \Pr[\, \bar{E}_1 \cap \bar{E}_2 \,] = 1 - \Pr[\, \bar{E}_1 \,] \Pr[\, \bar{E}_2 \,] = 1 - (1-p)^2 = 2p - p^2$$

The union bound (which doesn't require independence) gives

$$\Pr[\, E_1 \cup E_2 \,] \leq \Pr[\, E_1 \,] + \Pr[\, E_2 \,] = 2p$$

which is asymptotically the same when $p$ is small (the term $p^2$ is negligible), and much easier to derive.

For $k$ events, we similarly have

$$\Pr[\, E_1 \cup \cdots \cup E_k \,] = 1 - (1-p)^k = 1 - \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^\ell p^\ell = kp - \sum_{\ell=2}^{k} \binom{k}{\ell} (-1)^\ell p^\ell$$

which is quite cumbersome. But gain, even without independence the union bound directly gives the "pretty good bound"

$$\Pr[\, E_1 \cup \cdots \cup E_k \,] \leq kp$$

which will suffice most of the time.

**Problem 2.** Prove Chebyshev's inequality using Markov's inequality.

**Solution 2.** As seen during the lecture, we apply Markov's inequality to the non-negative random variable $Y = (X - \mathbb{E}[X])^2$: for every $t > 0$,

$$\Pr[\, |X - \mathbb{E}[X]| \geq t \,] = \Pr\Big[\, \sqrt{Y} \geq t \,\Big] = \Pr\Big[\, Y \geq t^2 \,\Big] \leq \frac{\mathbb{E}[Y]}{t^2} = \frac{\mathrm{Var}[X]}{t^2}$$

(Justify each step.)

**Problem 3.** Compute the expectation and variance of a Poisson($\lambda$) random variable. (Recall that if $X \sim \text{Poisson}(\lambda)$, then $\Pr[\, X = k \,] = e^{-\lambda} \frac{\lambda^k}{k!}$ for any integer $k \geq 0$.)

**Solution 3.** Manipulating the infinite sums. For instance, for the expectation: if $X \sim \text{Poisson}(\lambda)$, then

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k}{k!} = \lambda \cdot e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \cdot \underbrace{e^{-\lambda} \sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{\ell!}}_{=1} = \lambda$$

For the variance, compute $\mathbb{E}\left[X^2\right]$ similarly, then subtract $\mathbb{E}[X]^2 = \lambda^2$. The answer is again $\lambda$.

**Problem 4.** Let $X$ be a Binomial random variable with parameters $n$ and $p$. Compute (or recall) the expectation and variance of $X$.

a) Bound the probability that $X$ deviates from its expectation by more than $2\sqrt{np}$.

b) Suppose that $p = \frac{1}{4}$.
- Use Markov's inequality to bound $\Pr[\, X \geq n/2 \,]$.
- Use Chebyshev's inequality to bound $\Pr[\, X \geq n/2 \,]$.
- Use the Chernoff bound to bound $\Pr[\, X \geq n/2 \,]$.
- Use Hoeffding's bound to bound $\Pr[\, X \geq n/2 \,]$.
- Compare the 4 bounds.

c) Suppose now that $p = \frac{1}{2n}$.
- Use Markov's inequality to bound $\Pr[\, X \geq 1 \,]$.
- Use Chebyshev's inequality to bound $\Pr[\, X \geq 1 \,]$. Comment.
- Use the Chernoff bound to bound $\Pr[\, X \geq 1 \,]$.
- Use Hoeffding's bound to bound $\Pr[\, X \geq 1 \,]$.
- Compute $\Pr[\, X \geq 1 \,]$ exactly, and compare the bounds obtained.

**Solution 4.**

a) Since the variance of $X \sim \text{Bin}(n, p)$ is $\text{Var}[X] = np(1-p) \leq np$, Chebyshev's inequality gives $\Pr\left[\, |X - \mathbb{E}[X]| \geq 2\sqrt{np} \,\right] \leq \frac{np}{(2\sqrt{np})^2} = \frac{1}{4}$.

b)

- With Markov's, we get $\Pr[\, X \geq n/2 \,] \leq \frac{np}{(n/2)} = 2p = 1/2$
- The variance is $\mathrm{Var}[X] = np(1-p) = 3n/16$ and the expectation is $\mathbb{E}[X] = n/4$, so the bound we get via Chebyshev's is

$$\Pr[\, X \geq n/2 \,] = \Pr[\, X - n/4 \geq n/4 \,] \leq \Pr[\, |X - n/4| \geq n/4 \,] \leq \frac{3n/16}{(n/4)^2} = \frac{3}{n}$$

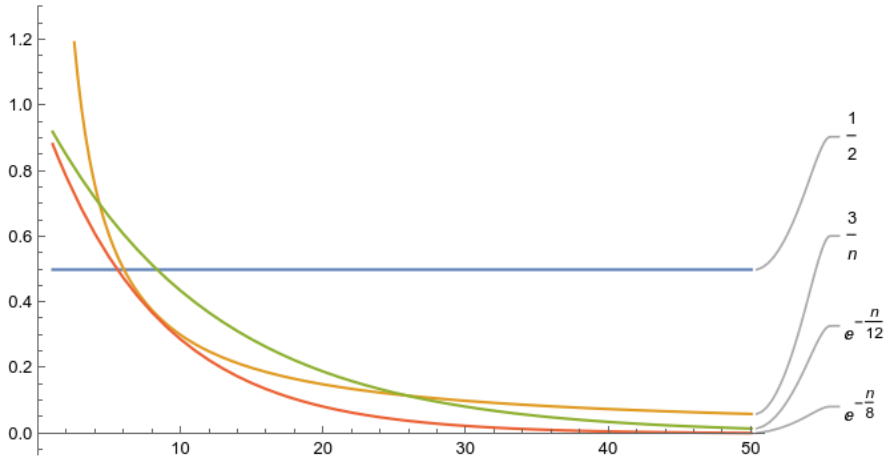
- We apply the Chernoff bound with $\gamma = 1$ and $P = \mathbb{E}[X] = n/4$:

$$\Pr\left[\, X \geq \frac{n}{2} \,\right] = \Pr[\, X \geq 2\mathbb{E}[X] \,] \leq e^{-\frac{\gamma^2 \mathbb{E}[X]}{3}} = e^{-\frac{n}{12}}$$

- We apply Hoeffding's bound with $t = n/4$, $\mathbb{E}[X] = n/4$, and $a_i = 0, b_i = 1$ for all $1 \leq i \leq n$:

$$\Pr\left[\, X \geq \frac{n}{2} \,\right] = \Pr[\, X \geq \mathbb{E}[X] + t \,] \leq e^{-\frac{2t^2}{n(1-0)^2}} = e^{-\frac{n}{8}}$$

- The bound obtained with Markov's inequality is the weakest (it does not even go to 0 as $n$ grows). Chebyshev's is better in that regard, as the bound obtained decreases with $n$ (technically, better than Markov's for $n \geq 6$): but it still only decreases polynomially. In contrast, both the Chernoff and Hoeffding bounds give upper bound that decrease exponentially fast with $n$, that is, of the form $e^{-\Theta(n)}$. In this particular setting, the result given by the Hoeffding's bound is slightly better than the Chernoff bound (the constant in the exponent is better).




That exponential decay is much stronger "for $n$ large enough", and is typically better: but due to the constants in the exponent, etc., it only really kicks in for $n \gg 1$.

c)

- Same result: $1/2$
- Chebyshev now yield a bound of $4(1 - 1/2n)$, which is a vacuous bound (more than 1 – probabilities are at most 1).
- $e^{-1/6} \approx 0.85$ (setting $\gamma = 1$, $P = 1/2$)
- $e^{-1/(2n)} \approx 1 - 1/(2n)$ (setting $t = 1/2$, $a_i = 0, b_i = 1$). Almost vacuous.
- Note that none of the bounds above decays to zero as $n$ grows... Which is maybe to be expected: the expectation itself is $1/2$, it looks like $n$ is not really "in the picture" here?
  We can compute the probability exactly as

$$\Pr[\, X \geq 1 \,] = 1 - \Pr[\, X = 0 \,] = 1 - (1 - 1/(2n))^n \approx 1 - e^{-1/2} \approx 0.39$$

  Maybe surprisingly, Markov is better here!

  *Side note:* in this parameter regime where the probability $p = p(n)$ is such that $\lim_{n\to\infty} np = \lambda$ for some constant $\lambda > 0$ (here $\lambda = 1/2$), a Binomial random variable with parameters $n, p$ "behaves like" (in some formal sense) a Poisson random variable with parameter $\lambda$. And you can check that if $X \sim \text{Poi}(\lambda)$, then $\Pr[\, X \geq 1 \,] = 1 - \Pr[\, X = 0 \,] = 1 - e^{-\lambda} = 1 - e^{-1/2} \approx 0.39$.

## Problem solving

**Problem 5.** Prove Theorem 8 of the lecture notes:

> Let $A$ be a Monte Carlo algorithm with worst-case running time $T(n)$ and constant failure probability $p \in (0, 1)$, with the following extra guarantee: one can detect whether the output of $A$ is incorrect in time $O(1)$. Then there exists a *Las Vegas* algorithm $A'$ for the same task with expected running time $O(T(n))$ (where the hidden constant in the $O(\cdot)$ depends on $p$).

**Solution 5.** See lecture video: we repeat a (new) call to $A$ and a check of the output until one iteration gives a correct output. Each repetition takes take time $T + O(1) = O(T)$, and the probability that we do at least $k \geq 0$ repetitions is the probability the first $k - 1$ failed to result in a good output, which happens with probability $p^{k-1}$. So we have

$$\mathbb{E}[\text{running time}] = \sum_{k=1}^{\infty} O(T) \cdot p^{k-1} = O(T) \cdot \sum_{k=0}^{\infty} p^k = O(T) \cdot \frac{1}{1-p} = O(T/(1-p)).$$

**Problem 6.** Suppose that we have two Monte Carlo algorithms $A$ and $B$ for a decision problem $P$, with the following behaviour: on any input $x$,

- if the true answer $P(x)$ is yes, then $A$ outputs yes with probability at least $1/2$, while $B$ outputs yes with probability one.
- if the true answer $P(x)$ is no, then $A$ outputs no with probability one, while $B$ outputs no with probability at least $1/2$.

Both $A$ and $B$ run in worst-case time $T(|x|)$. Using $A$ and $B$, design a Las Vegas algorithm $C$ for $P$. Analyse its expected running time.

**Solution 6.** The idea is, on input $x$, to run both $A$ and $B$ on $x$ (with independent random bits for each). If $A$ outputs "yes", then we know the true answer is "yes" (by the contrapositive of the second bullet); if $B$ outputs "no", we know the answer is "no" (by the contrapositive of the first bullet). If $A$ outputs "no" and $B$ outputs "yes", then we don't know for sure, so we repeat (until one of the two cases above occurs). We can check that

- If $P(x)$ is yes, then $\Pr[\, A$ outputs "no" and $B$ outputs "yes"$\,] \leq 1/2$.
- If $P(x)$ is no, then $\Pr[\, A$ outputs "no" and $B$ outputs "yes"$\,] \leq 1/2$.

So in ether case, the probability we have to repeat at a given step $k$ is $\leq 1/2$, and so, the expect running time is at most $(T(|x|) + T(|x|)) \cdot \sum_{k=0}^{\infty}(1/2)^k = O(T(|x|))$.

**Problem 7.** Let $A$ be a randomised algorithm which, on input $x$, consumes (at most) $T$ "resources" and uses (at most) $r$ random bits, outputs good or bad, such that

- If $x$ is good, then $\Pr[\, A(x) = \mathsf{good}\,] \geq 9/10$;
- If $x$ is bad, then $\Pr[\, A(x) = \mathsf{good}\,] \leq 1/10$.

For any $\delta \in (0, 1]$, give a randomised algorithm $A'$ such that, on input $x$,

- If $x$ is good, then $\Pr[\, A(x) = \mathsf{good}\,] \geq 1 - \delta$;
- If $x$ is bad, then $\Pr[\, A(x) = \mathsf{good}\,] \leq \delta$.

Bound the amount of resources $T'$ and random bits $r'$ this algorithm $A'$ uses.

**Solution 7.** Repeat $k$ times, for an integer $k = k(\delta)$ to be determined, and take a majority vote. Consider two approaches to analyse the number of repetitions: via Chebyshev, and via Hoeffding/Chernoff. Compare.

**Problem 8.** Similar, but a little different: Let $A$ be a randomised algorithm which, on input $x$, consumes (at most) $T$ "resources" and uses (at most) $r$ random bits, outputs good or bad, such that

- If $x$ is good, then $\Pr[\, A(x) = \mathsf{good}\,] \geq 1/10$;
- If $x$ is bad, then $\Pr[\, A(x) = \mathsf{good}\,] = 0$.

For any $\delta \in (0, 1]$, give a randomised algorithm $A'$ such that, on input $x$,

- If $x$ is good, then $\Pr[\, A(x) = \mathsf{good} \,] \geq 1 - \delta$;
- If $x$ is bad, then $\Pr[\, A(x) = \mathsf{good} \,] = 0$.

Bound the amount of resources $T'$ and random bits $r'$ this algorithm $A'$ uses.

**Solution 8.** Similar, but this time all we need to conclude is to see good be returned at least once (since this can *only* happen when $x$ is truly good): if it never happens in $k$ repetitions, we can return bad. The probability that good doesn't happen in $k$ repetitions when $x$ is good is at most $(1 - 1/10)^k = (9/10)^k$, so we only need to choose $k$ so that $(9/10)^k \leq \delta$, which leads to $k = \lceil \log_{10/9}(1/\delta) \rceil$.
*Note:* an algorithm which can only err in one of two cases is called an algorithm with *one-sided error.*

**Problem 9.** We will prove (a simplified version of) the Chernoff bound. Namely, given $X_1, \ldots, X_n$ i.i.d. random variables taking values in $\{0, 1\}$, each with expectation $p$, set $X = \sum_{i=1}^{n} X_i$. We will show that

$$\Pr[\, X > (1+\gamma)\mathbb{E}[X] \,] \leq e^{-\gamma^2 \mathbb{E}[X]/3}, \qquad \gamma \in (0, 1]$$

In what follows, fix any $\gamma \in (0, 1]$.

a) Show that, for every $t > 0$,

$$\Pr[\, X > (1+\gamma)\mathbb{E}[X] \,] = \Pr\left[\, e^{tX} > e^{t(1+\gamma)\mathbb{E}[X]} \,\right].$$

b) Deduce that, for every $t > 0$,

$$\Pr[\, X > (1+\gamma)\mathbb{E}[X] \,] \leq \frac{\mathbb{E}\left[e^{tX_1}\right]^n}{e^{t(1+\gamma)\mathbb{E}[X]}}.$$

c) Compute $\mathbb{E}\left[e^{tX_1}\right]$, and deduce that, for every $t > 0$,

$$\Pr[\, X > (1+\gamma)\mathbb{E}[X] \,] \leq \frac{(1 + p(e^t - 1))^n}{e^{t(1+\gamma)np}}.$$

d) Use the inequality $\ln(1 + x) \leq x$ to show that, for every $t > 0$,

$$\Pr[\, X > (1+\gamma)\mathbb{E}[X] \,] \leq e^{-pn \cdot f(t)}.$$

where $f(t) = (1+\gamma)t - (e^t - 1)$.

e) Choose the best value of $t > 0$ (which is a free parameter) to show that

$$\Pr[\, X > (1+\gamma)\mathbb{E}[X]\,] \leq e^{-pn((1+\gamma)\ln(1+\gamma)-\gamma)}\,.$$

Show (or take for granted, and verify by plotting the two functions) that $(1+\gamma)\ln(1+\gamma) - \gamma \geq \gamma^2/3$ for all $\gamma \in (0, 1]$. Conclude.

**Solution 9.**

a) This follows from recalling that $\Pr[\, X > Y\,] = \Pr[\, f(X) > f(Y)\,]$ for any increasing function $f$, and applying that to $f(x) = e^{tx}$. (Increasing for any fixed $t > 0$.)

b) First, Markov's inequality, then independence of $X_1, \ldots, X_n$ (and the fact they are identically distributed):

$$\Pr\left[\, e^{tX} > e^{t(1+\gamma)\mathbb{E}[X]}\,\right] \leq \frac{\mathbb{E}\left[e^{tX}\right]}{e^{t(1+\gamma)\mathbb{E}[X]}} = \frac{\mathbb{E}\left[e^{\sum_{i=1}^{n} tX_i}\right]}{e^{t(1+\gamma)\mathbb{E}[X]}} = \frac{\prod_{i=1}^{n}\mathbb{E}\left[e^{tX_i}\right]}{e^{t(1+\gamma)\mathbb{E}[X]}} = \frac{\mathbb{E}\left[e^{tX_1}\right]^n}{e^{t(1+\gamma)\mathbb{E}[X]}}$$

c) Since $X_1 \sim \text{Bern}(p)$,

$$\mathbb{E}\left[e^{tX_1}\right] = p \cdot e^{t\cdot 1} + (1-p)\cdot e^{t\cdot 0} = 1 + p(e^t - 1)$$

and for the denominator we also have $\mathbb{E}[X] = np$ by linearity of expectation. Note: when it exists, the function $\phi(t) = \mathbb{E}\left[e^{tX_1}\right]$ is called the *moment-generating function* (MGF) of $X_1$.

d) We have shown so far that (for any $t > 0$) $\Pr[\, X > (1+\gamma)\mathbb{E}[X]\,] \leq \frac{(1+p(e^t-1))^n}{e^{t(1+\gamma)np}}$ ..
Using the suggested inequality, we can bound the RHS further:

$$\frac{(1+p(e^t-1))^n}{e^{t(1+\gamma)np}} = \frac{e^{n\ln(1+p(e^t-1))}}{e^{t(1+\gamma)np}} \leq \frac{e^{np(e^t-1)}}{e^{t(1+\gamma)np}} = e^{np\left((e^t-1)-t(1+\gamma)\right)} = e^{-np\cdot f(t)}$$

where $f(t) = (1+\gamma)t - (e^t - 1)$, as stated in the question.
We introduced our "free parameter" $t > 0$ early on for a reason: the inequality

$$\Pr[\, X > (1+\gamma)\mathbb{E}[X]\,] \leq e^{-np\cdot f(t)}$$

holds for *every* choice of $t > 0$ (every choice of $t > 0$ we make gives a valid bound), so now we get to pick the best possible one to make the inequality as strong as possible (i.e., the RHS as small as possible). This corresponds to finding $t > 0$ maximizing $f(t)$, for which we can use calculus. $f$ is smooth, etc., so we can differentiate:

$$f'(t) = (1+\gamma) - e^t$$

and after solving $f'(t) = 0$ we can easily check that the critical point $t = \ln(1+\gamma) > 0$ is a maximum for $f$. Choosing this value of $t$ gives

$$\Pr[\, X > (1+\gamma)\mathbb{E}[X]\,] \leq e^{-np\cdot f(\ln(1+\gamma))} = e^{-np\cdot((1+\gamma)\ln(1+\gamma)-\gamma)}$$

Now, we can again show the suggested inequality

$$(1+\gamma)\ln(1+\gamma) - \gamma \geq \gamma^2/3, \qquad \gamma \in [0,1]$$

by calculus (studying the function $\gamma \mapsto \frac{(1+\gamma)\ln(1+\gamma)-\gamma}{\gamma^2}$, for instance); assuming this inequality, we get

$$\Pr[\, X > (1+\gamma)\mathbb{E}[X]\,] \leq e^{-np\cdot f(\ln(1+\gamma))} = e^{-np\cdot\frac{\gamma^2}{3}} = e^{-\frac{\gamma^2\mathbb{E}[X]}{3}}$$

and we are done. □
*Tip:* a Taylor expansion at $\gamma = 0$ of $(1+\gamma)\ln(1+\gamma) - \gamma$ gives

$$(1+\gamma)\ln(1+\gamma) - \gamma = \frac{\gamma^2}{2} - O(\gamma^3)$$

so the $\gamma^2$ dependence is tight (cannot be improved), and suggest why we might try to show this inequality in the first place. The constant 3 might be slightly improvable, but that's not critical.

## Advanced

**Problem 10.** Use the same approach to show the "other side" of the Chernoff bound:

$$\Pr[\, X < (1+\gamma)\mathbb{E}[X]\,] \leq e^{-\gamma^2 \mathbb{E}[X]/2}$$

for $\gamma \in (0,1]$. Do you see how to generalise the above argument to $X_1, \ldots, X_n \in [0,1]$? To independent (but non-identically distributed) $X_i's$?

**Problem 11.** We will prove the *median trick*. Suppose that any given input $x$ is associated with an interval $[a_x, b_x] \subseteq \mathbb{R}$ of "good values." We don't know this interval: our goal is, given any input $x$ to find a good value for $x$ with very high probability, say $1-\delta$ for arbitrarily small $\delta$.

All we are given is an algorithm $A$ which, on any input $x$, is guaranteed to output a good value with reasonably good probability. Specifically,

$$\Pr[\, A(x) < a_x\,] \leq \alpha, \qquad \Pr[\, A(x) > b_x\,] \leq \alpha$$

for some known constant $\alpha < 1/2$. Consider the following algorithm $B$: on input $x$, run $A$ on $x$ independently $k$ times, and output the median of all $k$ values obtained.

a) Analyse the probability that the output of $B$ is a good value, as a function of $\alpha$ and $k$.

b) Set the integer $k$ to achieve our original goal: output a good value with probability at least $1-\delta$.

**Solution 11.** *Sketch:* The key insight is that the median of $k$ outputs is less than $a_x$ if, and only if, more than $k/2$ of these values are less than $a_x$. But each of these $k$ things happens independently with probability at most $\alpha < 1/2$, so we can use either a Chernoff or Hoeffding bound to get that

$$\Pr\left[\text{ more than } \frac{k}{2} \text{values are less than } a_x \right] \leq e^{-\Theta(k)}$$

By setting $k \geq C \cdot \log(2/\delta)$ for a suitable constant $C > 0$, the RHS of that bound is at most $\delta/2$. We can do a similar analysis for the probability to be more than $b_x$, and then take a union bound over both events (this is why we chose $\delta/2$ above) to get that the probability the median is in $[a_x, b_x]$ is at least $1-\delta$.

*Note:* here we really had to use the fact that $\alpha < 1/2$ for it to work (can you see why?). We can do things a bit more directly and faster if our assumption is $\Pr[\, A(x) \notin [a_x, b_x]\,] \leq \alpha < 1/2$ (then we only have one event to handle for the median, not both + a union bound), but this is a stronger assumption (again, can you see why?).