HDP – Solutions

My solutions for the "High Dimensional Probability" (HDP) book. To be used for self-study. If you find mistakes and typos, or something is unclear, please let me know!

CHAPTER 0

Exercise 0.0.3

a) Expanding the left-hand side, we have:

$$\mathbb{E}\left\|\sum_{j=1}^{k}\overrightarrow{Z_j}\right\|_2^2 = \mathbb{E}\left[\left(\sum_{j=1}^{k}\overrightarrow{Z_j}\right)\cdot\left(\sum_{j=1}^{k}\overrightarrow{Z_j}\right)\right] = \mathbb{E}\left[\left(\sum_{j=1}^{k}\overrightarrow{Z_j}\right)\cdot\left(\sum_{j'=1}^{k}\overrightarrow{Z_{j'}}\right)\right].$$

Indeed, for any vector $\vec{v}$ we have that $\|\vec{v}\|_2^2 = \vec{v}\cdot\vec{v}$ (the dot product of a vector by itself is equal to the vector's length squared). Above, we applied this for $\vec{v} = \sum_{j=1}^{k}\overrightarrow{Z_j}$. Moreover, we can freely change the summation index in the 2nd sum. Continuing, we have:

$$\mathbb{E}\left[\left(\sum_{j=1}^{k}\overrightarrow{Z_j}\right)\cdot\left(\sum_{j'=1}^{k}\overrightarrow{Z_{j'}}\right)\right] = \mathbb{E}\left[\sum_{j=1}^{k}\sum_{j'=1}^{k}\overrightarrow{Z_j}\cdot\overrightarrow{Z_{j'}}\right] = \sum_{j=1}^{k}\sum_{j'=1}^{k}\mathbb{E}\left[\overrightarrow{Z_j}\cdot\overrightarrow{Z_{j'}}\right],$$

from the linearity of expectation. After that, since the vectors $\overrightarrow{Z_j}$ are assumed independent, we have that $\mathbb{E}\left[\overrightarrow{Z_j}\cdot\overrightarrow{Z_{j'}}\right] = \mathbb{E}\left[\overrightarrow{Z_j}\right]\cdot\mathbb{E}\left[\overrightarrow{Z_{j'}}\right]$ for $j \neq j'$. We split the summation in 2 terms, one sum where the indexes $j = j' = i$ are equal, and another, where they are different:

$$\sum_{j=1}^{k}\sum_{j'=1}^{k}\mathbb{E}\left[\overrightarrow{Z_j}\cdot\overrightarrow{Z_{j'}}\right] = \sum_{i=1}^{k}\mathbb{E}\left[\overrightarrow{Z_i}\cdot\overrightarrow{Z_i}\right] + \sum_{j,j'=1,j\neq j'}^{k}\mathbb{E}\left[\overrightarrow{Z_j}\cdot\overrightarrow{Z_{j'}}\right] =$$

$$= \sum_{i=1}^{k}\mathbb{E}\left[\left\|\overrightarrow{Z_i}\right\|_2^2\right] + \sum_{j,j'=1,j\neq j'}^{k}\mathbb{E}\left[\overrightarrow{Z_j}\right]\cdot\mathbb{E}\left[\overrightarrow{Z_{j'}}\right]$$

Since we assume that $\mathbb{E}\left[\overrightarrow{Z_j}\right] = \vec{0}$, the 2nd sum is zero, and we finally have:

$$\mathbb{E}\left\|\sum_{j=1}^{k}\overrightarrow{Z_j}\right\|_2^2 = \sum_{j=1}^{k}\mathbb{E}\left[\left\|\overrightarrow{Z_j}\right\|_2^2\right],$$

as required.

b) Set $\vec{\mu} = \mathbb{E}\vec{Z}$ (the mean of $\vec{Z}$). We have:

$$\mathbb{E}\left\|\vec{Z} - \mathbb{E}\vec{Z}\right\|_2^2 = \mathbb{E}\left\|\vec{Z} - \vec{\mu}\right\|_2^2 = \mathbb{E}\left[\left(\vec{Z} - \vec{\mu}\right) \cdot \left(\vec{Z} - \vec{\mu}\right)\right] = \mathbb{E}\left[\vec{Z} \cdot \vec{Z} - 2\vec{\mu} \cdot \vec{Z} + \vec{\mu} \cdot \vec{\mu}\right].$$

"Opening" the expectation, and remembering that $\mathbb{E}\left[\vec{\mu} \cdot \vec{Z}\right] = \vec{\mu} \cdot \mathbb{E}\left[\vec{Z}\right]$, since $\vec{\mu}$ is a constant vector, we find:

$$\mathbb{E}\left[\vec{Z} \cdot \vec{Z} - 2\vec{\mu} \cdot \vec{Z} + \vec{\mu} \cdot \vec{\mu}\right] = \mathbb{E}\left[\vec{Z} \cdot \vec{Z}\right] - 2\vec{\mu} \cdot \mathbb{E}\left[\vec{Z}\right] + \vec{\mu} \cdot \vec{\mu} = \mathbb{E}\left[\vec{Z} \cdot \vec{Z}\right] - 2\vec{\mu} \cdot \vec{\mu} + \vec{\mu} \cdot \vec{\mu} =$$

$$= \mathbb{E}\left[\vec{Z} \cdot \vec{Z}\right] - \vec{\mu} \cdot \vec{\mu}.$$

Finally, since $\mathbb{E}\left[\vec{Z} \cdot \vec{Z}\right] = \mathbb{E}\left\|\vec{Z}\right\|_2^2$ (as before) and putting back $\vec{\mu} = \mathbb{E}\vec{Z}$, we get:

$$\mathbb{E}\left[\vec{Z} \cdot \vec{Z}\right] - \vec{\mu} \cdot \vec{\mu} = \mathbb{E}\left\|\vec{Z}\right\|_2^2 - \mathbb{E}\vec{Z} \cdot \mathbb{E}\vec{Z} = \mathbb{E}\left\|\vec{Z}\right\|_2^2 - \left\|\mathbb{E}\vec{Z}\right\|_2^2,$$

as required.

Exercise 0.0.5

We will do each inequality separately. For the 1st, we have:

$$\binom{n}{m} = \frac{n(n-1)\ldots(n-m+1)}{m \cdot \ldots \cdot 2 \cdot 1} = \frac{n}{m} \cdot \frac{n-1}{m-1} \cdot \ldots \cdot \frac{n-m+1}{1}.$$

For a fraction a/b with $a \geq b$, we have that $\frac{a-c}{b-c} \geq \frac{a}{b}$ for $0 < c < b$. Indeed, multiplying out we have:

$$\frac{a-c}{b-c} \geq \frac{a}{b} \Leftrightarrow b(a-c) \geq a(b-c) \Leftrightarrow ba - bc \geq ab - ac \Leftrightarrow -bc \geq -ac \Leftrightarrow bc \leq ac \Leftrightarrow$$

$$\Leftrightarrow b \leq a,$$

which is true by assumption. In fact, since all steps are $\Leftrightarrow$, equality holds when $a = b$, and inequality when $a > b$. Using this property, we see that:

$$\frac{n}{m} \leq \frac{n-1}{m-1}, \frac{n}{m} \leq \frac{n-2}{m-2}, \ldots, \frac{n}{m} \leq \frac{n-(m-1)}{m-(m-1)} = \frac{n-m+1}{1}.$$

Therefore:

$$\binom{n}{m} = \frac{n(n-1)\ldots(n-m+1)}{m \cdot \ldots \cdot 2 \cdot 1} = \frac{n}{m} \cdot \frac{n-1}{m-1} \cdot \ldots \cdot \frac{n-m+1}{1} \geq \frac{n}{m} \cdot \ldots \cdot \frac{n}{m} = \left(\frac{n}{m}\right)^m,$$

which establishes the 1st inequality. Equality holds when $n = m$, because only in this case the fractions above are equal.

For the 2nd inequality, we have:

$$\binom{n}{m} \leq \sum_{k=0}^{m} \binom{n}{k}.$$

This holds, because the index k will also get the value m, and we will have the term $\binom{n}{m}$ on the right-hand side. The rest of the terms are all positive, so the right-hand side will be larger. The inequality is always strict for $n \geq 1$ and $m \in [1 \dots n]$.

Finally, we have the last inequality:

$$\sum_{k=0}^{m} \binom{n}{k} \leq \left(\frac{en}{m}\right)^m.$$

First, we multiply both sides by $\left(\frac{m}{n}\right)^m$, and we get:

$$\sum_{k=0}^{m} \binom{n}{k} \left(\frac{m}{n}\right)^m \leq e^m.$$

Since $\frac{m}{n} \leq 1$, we have that $\left(\frac{m}{n}\right)^m \leq \left(\frac{m}{n}\right)^k$, because $k \leq m$. So, replacing $\left(\frac{m}{n}\right)^m$ with $\left(\frac{m}{n}\right)^k$ on the right-hand side we get a larger quantity (with equality when $\frac{m}{n} = 1$), so it suffices to prove:

$$\sum_{k=0}^{m} \binom{n}{k} \left(\frac{m}{n}\right)^k \leq e^m.$$

But using the binomial theorem, the left-hand side is:

$$\sum_{k=0}^{m} \binom{n}{k} \left(\frac{m}{n}\right)^k = \sum_{k=0}^{m} \binom{n}{k} \left(\frac{m}{n}\right)^k (1)^{m-k} = \left(1 + \frac{m}{n}\right)^m,$$

hence:

$$\left(1 + \frac{m}{n}\right)^m \leq e^m \Leftrightarrow \left(1 + \frac{m}{n}\right) \leq e.$$

This is true, since $m \leq n \Leftrightarrow \frac{m}{n} \leq 1 \Leftrightarrow 1 + \frac{m}{n} \leq 2 < e = 2.718 \ldots$. Equality never holds.

Exercise 0.0.6

As in Corollary 0.0.4, we consider the set $\boldsymbol{N} = \left\{\frac{1}{k}\sum_{j=1}^{k} x_j : x_j \; vertices \; of \; P\right\}$. Then, any point within $conv(P)$ is within distance $d \leq 1/\sqrt{k}$ from some point in $\boldsymbol{N}$. To make the distance smaller or equal to ε, we can take:

$$\frac{1}{\sqrt{k}} \leq \varepsilon \Leftrightarrow k \geq \frac{1}{\varepsilon^2} \Rightarrow k = \left\lceil \frac{1}{\varepsilon^2} \right\rceil,$$

where $\lceil x \rceil$ is the ceil function (returning the 1st integer that is larger than x). This is needed, since k must be an integer.

Now, notice that the set $\boldsymbol{N}$ is the combination of k out of the total N vertices of P, and repetition is allowed. Thus, the cardinality (number of points) of $\boldsymbol{N}$ is the number of ways we can choose k out of the total N vertices with repetition. This number is $Crep(k, N) = \binom{N+k-1}{k}$.

From exercise 0.0.5, we have (with $n = N + k - 1, m = k$):

$$\binom{N+k-1}{k} \leq \left(\frac{e(N+k-1)}{k}\right)^k \leq \left(e\left(\frac{N+k}{k}\right)\right)^k = \left(e\left(1 + \frac{N}{k}\right)\right)^k.$$

We have that $k = \left\lceil \frac{1}{e^2} \right\rceil$, so with this we get:

$$|\boldsymbol{N}| \leq \left(e\left(1 + \frac{N}{k}\right)\right)^k = \left(e\left(1 + \frac{N}{\left\lceil \frac{1}{\varepsilon^2} \right\rceil}\right)\right)^{\left\lceil \frac{1}{\varepsilon^2} \right\rceil}.$$

Since $\left\lceil \frac{1}{\varepsilon^2} \right\rceil \geq \frac{1}{\varepsilon^2} \Leftrightarrow \frac{N}{\left\lceil \frac{1}{\varepsilon^2} \right\rceil} \leq \frac{N}{\frac{1}{\varepsilon^2}}$, so replacing $\left\lceil \frac{1}{\varepsilon^2} \right\rceil$ with $\frac{1}{\varepsilon^2}$ we get a larger quantity, so we have:

$$|\boldsymbol{N}| \leq \left(e\left(1 + \frac{N}{\frac{1}{\varepsilon^2}}\right)\right)^{\left\lceil \frac{1}{\varepsilon^2} \right\rceil} = \left(e(1 + \varepsilon^2 N)\right)^{\left\lceil \frac{1}{\varepsilon^2} \right\rceil} = (C + C\varepsilon^2 N)^{\left\lceil \frac{1}{\varepsilon^2} \right\rceil},$$

with $C = e$ (a constant). This is what we wanted to show.

CHAPTER 1

Exercise 1.2.2
Similarly as in the book, for every real x we have:

$$x = \int_0^\infty \mathbb{1}(x > t)dt - \int_{-\infty}^0 \mathbb{1}(x < t)dt.$$

We can see this by taking cases:

- If $x > 0$, then the 2nd integral is 0, and we get $x = \int_0^\infty \mathbb{1}(x > t)dt = \int_0^x 1dt = x.$
- If $x < 0$, the 1st integral is 0, so we get: $x = -\int_{-\infty}^0 \mathbb{1}(x < t)dt = -\int_x^0 1dt = \int_0^x 1dt = x$

Taking expectations on both sides, we get:

$$\mathbb{E}x = \int_0^\infty \mathbb{E}\mathbb{1}(x > t)dt - \int_{-\infty}^0 \mathbb{E}\mathbb{1}(x < t)dt.$$

But $\mathbb{E}\mathbb{1}(x > t) = \int_{-\infty}^\infty \mathbb{1}(x > t)p(x)dx = \int_t^\infty \mathbb{1}(x > t)p(x)dx = P(x > t)$, because the indicator function vanishes for $x \leq t$. In the same way, we find that $\mathbb{E}\mathbb{1}(x < t) = P(X < t)$. These finally give:

$$\mathbb{E}x = \int_0^\infty P(X > t)dt - \int_{-\infty}^0 P(X < t)dt.$$

Exercise 1.2.3
Using the integral identity from Lemma 1.2.1, we have:

$$|X|^p = \int_0^\infty \mathbb{1}(|X|^p > t)dt.$$

Set $t = u^p$, hence $dt = pu^{p-1}du$, and for the integration limits we have $u = 0$ for $t = 0$ and $u = \infty$ for $t = \infty$. Making this substitution, we have:

$$|X|^p = \int_0^\infty \mathbb{1}(|X|^p > u^p)pu^{p-1}du = \int_0^\infty pu^{p-1}\mathbb{1}(|X| > u)du.$$

Taking expectations, this yields:

$$\mathbb{E}|X|^p = \int_0^\infty pu^{p-1}E\mathbb{1}(|X| > u)du = \int_0^\infty pu^{p-1}P(|X| > u)du,$$

since $E\mathbb{1}(|X| > u) = \int_{-\infty}^{\infty} \mathbb{1}(|x| > u)p(x)dx = \int_{u}^{\infty} \mathbb{1}(x > u)p(x)dx + \int_{-\infty}^{u} \mathbb{1}(x > u)p(x)dx$
$= P(X \in [u, \infty]) + P(X \in [-\infty, u]) = P(|X| > u).$

This is what we wanted to show.

Exercise 1.2.6

For the random variable $U = (X - \mu)^2$, we have: $\mathbb{E}U = \mathbb{E}(X - \mu)^2 = Var[X] = \sigma^2$. Applying Markov's inequality on U, we get:

$$P(U \geq t^2) \leq \frac{\mathbb{E}U}{t^2} = \frac{\sigma^2}{t^2}.$$

But it holds that: $P(U \geq t^2) = P((X - \mu)^2 \geq t^2) = P(|X - \mu| \geq t)$, so finally:

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2},$$

as required. This is Chebyshev's inequality.

Exercise 1.3.3

We can write:

$$\frac{1}{N}\sum_{i=1}^{N} X_i - \mu = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu),$$

since $\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu) = \frac{1}{N}\sum_{i=1}^{N} X_i - \frac{1}{N}N\mu = \frac{1}{N}\sum_{i=1}^{N} X_i - \mu.$

Taking the variance, and noticing that X_i are independent, we get:

$$Var\left[\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)\right] = \frac{1}{N^2}\sum_{i=1}^{N} Var[(X_i - \mu)] = \frac{1}{N^2}\sum_{i=1}^{N} Var[X_i] = \frac{1}{N^2}N\sigma^2 = \frac{\sigma^2}{N}.$$

Now, we notice that for a random variable with mean 0, we have:

$$STD(X) = \sqrt{\mathbb{E}(X - \mu)^2} = \sqrt{\mathbb{E}X^2} \geq \mathbb{E}\sqrt{X^2} = \mathbb{E}|X|.$$

This follows from Jensen's inequality: indeed, since the function $f(x) = \sqrt{x}$ is concave ($f''(x) = \frac{3}{4x^{-\frac{3}{2}}} > 0$ for $x > 0$), Jensen's inequality gives: $f(\mathbb{E}U) \geq \mathbb{E}f(U)$ for every positive random variable U. Above, we applied this for $U = X^2$ and got:

$$\sqrt{\mathbb{E}X^2} = f(\mathbb{E}X^2) \geq \mathbb{E}f(X^2) = \mathbb{E}\sqrt{X^2} = \mathbb{E}|X|.$$

This is what we wanted to show.

CHAPTER 2

Exercise 2.1.4

We have (since $g \sim N(0,1)$):

$$\mathbb{E}[g^2 \mathbb{1}(g > t)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} \mathbb{1}(x > t) dx = \frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} x^2 e^{-x^2/2} dx = J,$$

since $\mathbb{1}(x > t)$ is zero in the interval $(-\infty, t)$. Consider now the following integral:

$$I = \frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} 1 e^{-x^2/2} dx = P(x > t).$$

We can set$1 = x'$ and integrate by parts. Doing this, we get:

$$I = \frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} (x)' e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} x e^{-x^2/2} \Big|_{t}^{\infty} - \frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} x \left(e^{-\frac{x^2}{2}} \right)' dx \Leftrightarrow$$

$$I = \left(0 - \frac{1}{\sqrt{2\pi}} t e^{-\frac{t^2}{2}} \right) - \frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} x \left(-x e^{-\frac{x^2}{2}} \right) dx \Leftrightarrow$$

$$I = -\frac{1}{\sqrt{2\pi}} t e^{-\frac{t^2}{2}} - \frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} x \left(-x e^{-\frac{x^2}{2}} \right) dx \Leftrightarrow$$

$$I = -\frac{1}{\sqrt{2\pi}} t e^{-\frac{t^2}{2}} + \frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} x^2 e^{-x^2/2} dx \Leftrightarrow$$

$$J = \frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} x^2 e^{-x^2/2} dx = I + \frac{1}{\sqrt{2\pi}} t e^{-\frac{t^2}{2}} = P(x > t) + t \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \Leftrightarrow$$

$$\frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} x^2 e^{-x^2/2} dx = P(x > t) + t \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}},$$

as required. This is the equality part, and holds for all t. Now, for the inequality, we can use proposition 2.1.2 from the text that shows $P(x \geq t) \leq \frac{1}{t}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$, so substituting this above, we get:

$$\frac{1}{\sqrt{2\pi}}\int_t^\infty x^2 e^{-x^2/2}dx = P(x > t) + t\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}} \leq \left(\frac{1}{t} + t\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}},$$

which completes the proof.

Exercise 2.2.3

For the Taylor expansion for $\cosh(x)$ we have:

$$\cosh(x) = \frac{e^x + e^{-x}}{2} = \frac{1}{2}\left[\left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} \dots\right) + \left(1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \cdots\right)\right] \Leftrightarrow$$

$$\cosh(x) = \frac{1}{2}\left[2\left(1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots\right)\right] = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots = \sum_{i=0}^{\infty}\frac{x^{2i}}{(2i)!}.$$

This is valid for all $x \in \mathbb{R}$. On the other hand, for the expansion of $\exp\left(\frac{x^2}{2}\right)$ we get:

$$\exp\left(\frac{x^2}{2}\right) = 1 + \frac{x^2}{2} + \frac{\left(\frac{x^2}{2}\right)^2}{2!} + \frac{\left(\frac{x^2}{2}\right)^3}{3!} + \cdots = \sum_{i=0}^{\infty}\frac{\left(\frac{x^2}{2}\right)^i}{i!} = \sum_{i=0}^{\infty}\frac{x^{2i}}{2^i i!}.$$

But we have that $2^i i! = i! \cdot 2 \cdot \dots \cdot 2 \leq i! \cdot (i+1) \cdot \dots \cdot (2i)$, and therefore

$$\frac{x^{2i}}{(2i)!} \leq \frac{x^{2i}}{2^i i!}$$

for all terms. Therefore, $\sum_{i=0}^{\infty}\frac{x^{2i}}{(2i)!} \leq \sum_{i=0}^{\infty}\frac{x^{2i}}{2^i i!} \Leftrightarrow \cosh(x) \leq \exp\left(\frac{x^2}{2}\right)$.

Exercise 2.2.7

In the beginning, we follow the same steps as in the proof of Theorem 2.2.2: namely, we multiply by λ and exponentiate. This gives:

$$P\left(\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \geq t\right) = P\left(\lambda\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \geq \lambda t\right) =$$

$$= P\left(epx\left(\lambda\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)\right) \geq \exp(\lambda t)\right). \qquad (2.2.6a)$$

Next, we apply Markov's inequality, as in Theorem 2.2.2:

$$P\left(\exp\left(\lambda\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)\right) \geq \exp(\lambda t)\right) \leq e^{-\lambda t}\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)\right)\right].$$

Now, using independence we have:

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)\right)\right] = \prod_{i=1}^{n}\mathbb{E}\big[\exp\big(\lambda(X_i - \mathbb{E}X_i)\big)\big]. \qquad (2.2.6b)$$

From here, we will try to bound every term in the above product. To do this, we will need Hoeffding's Lemma:

Hoeffding's Lemma: Let $X \in [a,b]$ be a random variable with $\mathbb{E}[X] = 0$, $a < b$. Then, for every $t > 0$ we have: $\mathbb{E}[e^{tX}] \leq e^{\frac{t^2(b-a)^2}{8}}$.

Proof: Since $f(x) = e^{tx}$ is convex ($(e^{tx})'' = t^2e^{tx} > 0, \forall x \in \mathbb{R}$) we have that for all $u, v \in [a,b]$ and $0 \leq \lambda \leq 1$ that:

$$f(\lambda u + (1-\lambda)v) \leq \lambda f(u) + (1-\lambda)f(v).$$

In the above, we select $x = a, y = b$ and $\lambda = \frac{b-x}{b-a}$ for some $x \in [a,b]$. Indeed, since $a \leq x$, we have that $b - x \leq b - a \Rightarrow \lambda = \frac{b-x}{b-a} \leq 1$, and also $\lambda \geq 0$ since $x \leq b$. With these, we have also that $1 - \lambda = 1 - \frac{b-x}{b-a} = \frac{(b-a)-(b-x)}{b-a} = \frac{x-a}{b-a}$, and $\lambda a + (1-\lambda)b = \frac{b-x}{b-a}a + \frac{x-a}{b-a}b = \frac{x(b-a)}{b-a} = x$, which means that $f(\lambda a + (1-\lambda)b) = f(x) = e^{tx}$.
Plugging in, we get:

$$e^{tx} \leq \frac{b-x}{b-a}e^{ta} + \frac{x-a}{b-a}e^{tb}.$$

Therefore:

$$\mathbb{E}[e^{tX}] \leq \mathbb{E}\left[\frac{b-x}{b-a}e^{ta} + \frac{x-a}{b-a}e^{tb}\right] = \frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb} = e^{\varphi(t)},$$

where we define the function

$$\varphi(t) = \ln\left(\frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb}\right) = \ln\left(e^{ta}\left(\frac{b}{b-a} - \frac{a}{b-a}e^{t(b-a)}\right)\right) =$$

$$= \ln(e^{ta}) + \ln\left(\frac{b}{b-a} - \frac{a}{b-a}e^{t(b-a)}\right) = ta + \ln\left(\frac{b}{b-a} - \frac{a}{b-a}e^{t(b-a)}\right).$$

Taking derivatives, we have:

$$\varphi'(t) = a + \frac{1}{\left(\frac{b}{b-a} - \frac{a}{b-a}e^{t(b-a)}\right)}\left(\frac{b}{b-a} - \frac{a}{b-a}e^{t(b-a)}\right)' =$$

$$= a + \frac{1}{\left(\frac{b}{b-a} - \frac{a}{b-a}e^{t(b-a)}\right)}\left(-\frac{a}{b-a}e^{t(b-a)}(b-a)\right) =$$

$$= a - \frac{ae^{t(b-a)}}{\left(\frac{b}{b-a} - \frac{a}{b-a}e^{t(b-a)}\right)} = a - \frac{a}{\left(\frac{b}{b-a}e^{-t(b-a)} - \frac{a}{b-a}\right)}.$$

First, we have that

$$\varphi(t) = \ln\left(\frac{b}{b-a}e^{0} - \frac{a}{b-a}e^{0}\right) = \ln\left(\frac{b-a}{b-a}\right) = \ln(1) = 0,$$

$$\varphi'(0) = a - \frac{a}{\left(\frac{b}{b-a}e^{0} - \frac{a}{b-a}\right)} = a - \frac{a}{\frac{b-a}{b-a}} = a - a = 0.$$

Moreover, we have:

$$\varphi''(t) = -\left(-\frac{a}{\left(\frac{b}{b-a}e^{-t(b-a)} - \frac{a}{b-a}\right)^2}\left(\frac{b}{b-a}e^{-t(b-a)} - \frac{a}{b-a}\right)'\right) =$$

$$= \frac{a}{\left(\frac{b}{b-a}e^{-t(b-a)} - \frac{a}{b-a}\right)^2}\left(-\frac{b}{b-a}e^{-t(b-a)}(b-a)\right) =$$

$$= -\frac{abe^{-t(b-a)}}{\left(\frac{b}{b-a}e^{-t(b-a)} - \frac{a}{b-a}\right)^2}.$$

To continue, some algebra is required. We set $c = -\frac{a}{b-a}$, so $1 - c = \frac{b}{b-a}$ and $c(1-c) = -ab/(b-a)^2$. Substituting c, we get:

$$\varphi''(t) = -\frac{abe^{-t(b-a)}}{\left(\frac{b}{b-a}e^{-t(b-a)} - \frac{a}{b-a}\right)^2} = \frac{c(1-c)e^{-t(b-a)}}{\left((1-c)e^{-t(b-a)} + c\right)^2}(b-a)^2 =$$

$$= \frac{c}{(1-c)e^{-t(b-a)} + c} \cdot \frac{(1-c)e^{-t(b-a)}}{(1-c)e^{-t(b-a)} + c}(b-a)^2.$$

Then, set $u = \frac{c}{(1-c)e^{-t(b-a)}+c} \Leftrightarrow 1 - u = \frac{(1-c)e^{-t(b-a)}}{(1-c)e^{-t(b-a)}+c}$, therefore:

$$\varphi''(t) = u(1-u)(b-a)^2. \tag{2.2.6c}$$

Consider now the function $g(u) = u(1-u) = u - u^2$. We gave $g'(u)1 - 2u \Rightarrow g'(u) = 0 \Leftrightarrow u = \frac{1}{2}$. Moreover, $g''(u) = -2 < 0$. Therefore, the critical point $u = \frac{1}{2}$ is a maximum, and we have:

$$g(u) \leq g\left(\frac{1}{2}\right) = \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4}.$$

Plugging this back into $(2.2.6c)$, we have $\varphi''(t) = (b-a)^2/4$. Finally, from the (extended) Mean Value Theorem, there must exist a $0 \leq \theta \leq t$ such that:

$$\varphi(t) = \varphi(0) + t\varphi'(0) + \frac{t^2}{2}\varphi''(\theta).$$

But $\varphi(0) = \varphi'(0) = 0$, and $\varphi''(\theta) \leq (b-a)^2/4$, so finally we get:

$$\varphi(t) \leq \frac{t^2}{8}(b-a)^2 \Rightarrow \mathbb{E}[e^{tX}] = e^{\varphi(t)} \leq e^{\frac{t^2}{8}(b-a)^2} \Rightarrow$$

$$\boxed{\mathbb{E}[e^{tX}] \leq e^{\frac{t^2}{8}(b-a)^2}}.$$

This proves Hoeffding's Lemma. □

After all this, we go back at eq. $(2.2.6b)$ and apply Hoeffding's Lemma on the random variables $Y_i = X_i - \mathbb{E}X_i$. We have that $\mathbb{E}[Y_i] = 0$, and since X_i lies in the interval $[m_i, M_i]$, Y_i lies in the interval $[y_i, Y_i] = [m_i - \mathbb{E}X_i, M_i - \mathbb{E}X_i]$. With these, Hoeffding's Lemma can be applied, and gives:

$$\mathbb{E}\left[\exp\left(\lambda(X_i - \mathbb{E}X_i)\right)\right] \leq e^{\frac{\lambda^2}{8}\left(M_i - \mathbb{E}X_i - (m_i - \mathbb{E}X_i)\right)^2} = e^{\frac{\lambda^2}{8}(M_i - m_i)^2}.$$

Putting this back in $(2.2.6b)$, we get:

$$P\left(\exp\left(\lambda\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)\right) \geq \exp(\lambda t)\right) \leq e^{-\lambda t}\prod_{i=1}^{n}\mathbb{E}\left[\exp\left(\lambda(X_i - \mathbb{E}X_i)\right)\right] =$$

$$= e^{-\lambda t}\prod_{i=1}^{n} e^{\frac{\lambda^2}{8}(M_i - m_i)^2} = \exp\left(-\lambda t + \frac{\lambda^2}{8}\sum_{i=1}^{n}(M_i - m_i)^2\right),$$

and finally, from $(2.2.6a)$ we have:

$$P\left(\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \geq t\right) \leq \exp\left(-\lambda t + \frac{\lambda^2}{8}\sum_{i=1}^{n}(M_i - m_i)^2\right). \qquad (2.2.6d)$$

This holds for all $\lambda > 0$, so we can minimize it with respect to λ. For this, set $\sum_{i=1}^{n}(M_i - m_i)^2 = A$ and consider the function $h(\lambda) = \exp\left(-\lambda t + \frac{\lambda^2}{8}A\right)$. We have:

$$h'(\lambda) = \exp\left(-\lambda t + \frac{\lambda^2}{8}A\right)' = e^{-\lambda t + \frac{\lambda^2}{8}A}\left(-t + \frac{\lambda A}{4}\right),$$

$$h'(\lambda) = 0 \Leftrightarrow \lambda^* = 4t/A.$$

$\lambda^* = 4t/A$ is a minimum, since $h'(\lambda)$ is negative for $\lambda < \lambda^*$ and positive for $\lambda > \lambda^*$, so $h(\lambda)$ is strictly decreasing before λ^*, and strictly increasing afterwards. So, λ^* is indeed the minimum.

Inserting this in $(2.2.6d)$ gives:

$$P\left(\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \geq t\right) \leq \exp\left(-\left(\frac{4t}{A}\right)t + \frac{\left(\frac{4t}{A}\right)^2 A}{8}\right) = \exp(-2t^2/A) \Leftrightarrow$$

$$\boxed{P\left(\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(M_i - m_i)^2}\right)}.$$

This (finally!) completes the proof. □

Exercise 2.2.8

Consider the indicator variable X_i, which is 1 if the algorithm A answers i-th trial correctly, and -1 otherwise. We will fail if the majority vote is failure, e.g. the sum of the X_i's is positive. Thus, the failure probability is:

$$P_{fail} = P\left(\sum_{i=1}^{n} X_i \geq 0\right).$$

The random variables X_i lie in the interval $[-1,1]$, and furthermore we have:

$$\mathbb{E}X_i = 1p - 1(1-p) = 1\left(\frac{1}{2} - \delta\right) - 1\left(\frac{1}{2} + \delta\right) = -2\delta,$$

since the failure probability is given as $p = \frac{1}{2} - \delta$. Since the X_i's are bounded but their probabilities are not symmetric, we have to use Theorem 2.2.6 in order to bound $P(\sum_{i=1}^{n} X_i \geq 0)$. But to do this, we have to also "bring in" the expectations $\mathbb{E}X_i$. For that, notice that the condition $\sum_{i=1}^{n} X_i \geq 0$ is equivalent to the following:

$$\sum_{i=1}^{n} X_i \geq 0 \Leftrightarrow \sum_{i=1}^{n} (X_i - \mathbb{E}\, X_i) \geq 0 - \sum_{i=1}^{n} \mathbb{E}X_i = 0 - \sum_{i=1}^{n} (-2\delta) = 2n\delta,$$

since $\mathbb{E}X_i = -2\delta$. Now we are ready to apply Hoeffding's theorem, and we have:

$$P_{fail} = P\left(\sum_{i=1}^{n} X_i \geq 0\right) = P\left(\sum_{i=1}^{n} (X_i - \mathbb{E}X_i) \geq 2n\delta\right) \leq \exp\left(-\frac{2(2n\delta)^2}{\sum_{i=1}^{n}(M_i - m_i)^2}\right) \Rightarrow$$

$$P_{fail} \leq \exp\left(-\frac{2(2n\delta)^2}{\sum_{i=1}^{n}\left(1-(-1)\right)^2}\right) = \exp\left(-\frac{2(2n\delta)^2}{4n}\right) = \exp(-2n\delta^2).$$

We want that the failure probability is upper bounded by ε. We can achieve that by selecting a n such that $\exp(-2n\delta^2) \leq \varepsilon$. Solving for n we get:

$$\exp(-2n\delta^2) \leq \varepsilon \Leftrightarrow -2n\delta^2 \leq \ln(\varepsilon) \Leftrightarrow n \geq \frac{(-\ln(\varepsilon))}{2\delta^2} \Leftrightarrow$$

$$\boxed{n \geq \frac{1}{2\delta^2} \ln\left(\frac{1}{\varepsilon}\right)}.$$

This is the number of trials (runs) which guarantees that the overall failure probability will be smaller than ε. □

Exercise 2.2.9

(a) We follow the hint and select the sample mean $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{n} X_i$ as the estimator. We want an epsilon accuracy with probability at least $1 - \delta$. This means, that we would like to have:

$$P(|\mu - \hat{\mu}| \geq \varepsilon) \leq \delta.$$

That is, we want the failure probability ($\hat{\mu}$ exceeds μ by more than ε) to be bounded by δ. Consider the random variable $Z = \frac{1}{N}\sum_{i=1}^{n} X_i$. We have $\mathbb{E}Z = \frac{1}{N}\sum_{i=1}^{n} \mathbb{E}X_i = \frac{1}{N}\sum_{i=1}^{n} \mu = \mu$, and

$$Var[Z] = \frac{1}{N^2}\sum_{i=1}^{n} Var[X_i] = \frac{1}{N^2}\sum_{i=1}^{n} \sigma^2 = \frac{1}{N^2} N\sigma^2 = \frac{\sigma^2}{N}.$$

Now, we can apply Chebyshev's inequality:

$$P(|\mu - \hat{\mu}| \geq \varepsilon) = P(|Z - \mu| \geq \varepsilon) \leq \frac{Var[Z]}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 N}.$$

We want to bound this by δ, so it suffices to select an N such that:

$$\frac{\sigma^2}{\varepsilon^2 N} \leq \delta \Leftrightarrow N \geq \frac{\sigma^2}{\varepsilon^2 \delta}.$$

In our case, $\delta = 1/4$, hence $\boxed{N \geq \frac{\sigma^2}{\varepsilon^2\left(\frac{1}{4}\right)} = 4\frac{\sigma^2}{\varepsilon^2} = O\left(\frac{\sigma^2}{\varepsilon^2}\right)}$ as required. □

(b) Consider k estimators as in part (a) and take their median. This will fail only if the media falls outside the interval $[\mu - \varepsilon, \mu + \varepsilon]$, and this will happen if and only if at least half of the estimators falls below $\mu - \varepsilon$ or at least half of the estimators fall above $\mu + \varepsilon$. Notice that this event is included in the event "at least half of the estimators fail" (because the later event includes the cases of the former, as well as cases where some estimators fall below $\mu - \varepsilon$ and some others above $\mu + \varepsilon$ – in some of these cases the median might be still inside $[\mu - \varepsilon, \mu + \varepsilon]$).

Consider now the random variables $Y_j, j = 1,2, \ldots, k$, where $Y_j = 1$ if the i-th estimator fails (e.g. falls outside $[\mu - \varepsilon, \mu + \varepsilon]$), and 0 otherwise. We want that at least half of the estimators fails, hence, the overall failure probability is:

$$P_{fail} \leq P\left(\sum_{j=1}^{k} Y_j \geq \frac{k}{2}\right).$$

This looks like a case for theorem 2.2.6, since the Y_j's are bounded. But to apply 2.2.6 we have to "bring in" the means $\mathbb{E}Y_j$, so we rewrite the condition above as:

$$\sum_{j=1}^{k} Y_j \geq \frac{k}{2} \Leftrightarrow \sum_{j=1}^{k} \left(Y_j - \mathbb{E}Y_j\right) \geq \frac{k}{2} - \sum_{j=1}^{k} \mathbb{E}Y_j.$$

The failure probability of Y_j is at most $1/4$, hence $\mathbb{E}Y_j \leq 1/4$, and thus

$$\sum_{j=1}^{k} \mathbb{E}Y_j \leq \frac{k}{4} \Leftrightarrow -\sum_{j=1}^{k} \mathbb{E}Y_j \geq -\frac{k}{4} \Leftrightarrow \frac{k}{2} - \sum_{j=1}^{k} \mathbb{E}Y_j \geq \frac{k}{4}.$$

Hence,

$$\sum_{j=1}^{k} \left(Y_j - \mathbb{E}Y_j\right) \geq \frac{k}{2} - \sum_{j=1}^{k} \mathbb{E}Y_j \geq \frac{k}{4},$$

hence the event $\sum_{j=1}^{k}\left(Y_j - \mathbb{E}Y_j\right) \geq \frac{k}{4}$ includes the event $\sum_{j=1}^{k}\left(Y_j - \mathbb{E}Y_j\right) \geq \frac{k}{2} - \sum_{j=1}^{k} \mathbb{E}Y_j$. So, $P\left(\sum_{j=1}^{k}\left(Y_j - \mathbb{E}Y_j\right) \geq \frac{k}{4}\right) \geq P\left(\sum_{j=1}^{k}\left(Y_j - \mathbb{E}Y_j\right) \geq \frac{k}{2} - \sum_{j=1}^{k} \mathbb{E}Y_j\right)$, and thus it suffices to bound the probability $P\left(\sum_{j=1}^{k}\left(Y_j - \mathbb{E}Y_j\right) \geq \frac{k}{4}\right)$. That is, we have:

$$P_{fail} \leq P\left(\sum_{j=1}^{k} Y_j \geq \frac{k}{2}\right) = P\left(\sum_{j=1}^{k}\left(Y_j - \mathbb{E}Y_j\right) \geq \frac{k}{2} - \sum_{j=1}^{k} \mathbb{E}Y_j\right) \leq P\left(\sum_{j=1}^{k}\left(Y_j - \mathbb{E}Y_j\right) \geq \frac{k}{4}\right).$$

Now, this has the form for Heffding's theorem 2.2.6, and we get:

$$P_{fail} \leq P\left(\sum_{j=1}^{k}\left(Y_j - \mathbb{E}Y_j\right) \geq \frac{k}{4}\right) \leq \exp\left(-\frac{2\left(\frac{k}{4}\right)^2}{\sum_{i=1}^{n}(M_i - m_i)^2}\right).$$

For the Y_j's, $m_i = 0, M_i = 1$, and bounding the above quantity by δ, we have:

$$P_{fail} \leq \exp\left(-\frac{2\left(\frac{k}{4}\right)^2}{\sum_{i=1}^{n}(M_i - m_i)^2}\right) = \exp\left(-\frac{2\left(\frac{k}{4}\right)^2}{\sum_{i=1}^{n} 1^2}\right) \leq \delta \Rightarrow$$

$$\exp\left(-\frac{k^2/8}{k}\right) = \exp\left(-\frac{k}{8}\right) \leq \delta \Rightarrow$$

$$-\frac{k}{8} \leq \ln(\delta) \Leftrightarrow k \geq 8 \ln\left(\frac{1}{\delta}\right).$$

Thus, with $k \geq 8\ln\left(\frac{1}{\delta}\right)$ estimators we can bound the failure probability by δ, hence $k = O\left(\ln\left(\frac{1}{\delta}\right)\right)$ estimators suffice. Since each estimator uses $O\left(\frac{\sigma^2}{\varepsilon^2}\right)$ samples, the total number of samples is $N = O\left(\ln\left(\frac{1}{\delta}\right)\frac{\sigma^2}{\varepsilon^2}\right)$. □

Exercise 2.2.10

a) Since X_i are non-negative, and their density is bounded by 1, for the MGF we have:

$$\mathbb{E}e^{-tX_i} = \int_{-\infty}^{\infty} e^{-tx}p(x)dx = \int_{0}^{\infty} e^{-tx}p(x)dx \leq \int_{0}^{\infty} e^{-tx}1dx = -\frac{e^{-tx}}{t}\bigg|_0^\infty = \frac{1}{t},$$

as required. The equality is because X_i is non-negative, and the inequality because $p(x) \leq 1$, where $p(x)$ is the probability density of X_i.

b) The given inequality can be written as:

$$\sum_{i=1}^{n} X_i \leq \varepsilon n \Leftrightarrow \sum_{i=1}^{n} \frac{X_i}{\varepsilon} \leq n \Leftrightarrow \sum_{i=1}^{n} \frac{-X_i}{\varepsilon} \geq -n.$$

Now, we proceed as in the proof of Hoedffding's inequality in the book, and arrive to the analogue of equation 2.5, where $a_i = -\frac{1}{\varepsilon}, t = -n$. This gives:

$$P_0 = P\left(\sum_{i=1}^{n} X_i \leq \varepsilon n\right) = P\left(\sum_{i=1}^{n} \frac{-X_i}{\varepsilon} \geq -n\right) \leq e^{-\lambda(-n)}\mathbb{E}\exp\left(\lambda \sum_{i=1}^{n} \frac{-X_i}{\varepsilon}\right) = e^{\lambda n}\prod_{i=1}^{n} \mathbb{E}e^{-\frac{\lambda}{\varepsilon}X_i}.$$

Applying part (a) on $\mathbb{E}e^{-\frac{\lambda}{\varepsilon}X_i}$(with $t = \frac{\lambda}{\varepsilon}$), we have that $\mathbb{E}e^{-\frac{\lambda}{\varepsilon}X_i} \leq 1/(\lambda/\varepsilon) = \varepsilon/\lambda$. Plugging this above, we get:

$$P_0 \leq e^{\lambda n}\prod_{i=1}^{n} \frac{\varepsilon}{\lambda} = e^{\lambda n}\frac{\varepsilon^n}{\lambda^n}.$$

Setting $\lambda = 1$ (since the above holds for all $\lambda > 0$), we get the desired result:

$$P_0 = P\left(\sum_{i=1}^{n} X_i \leq \varepsilon n\right) \leq \frac{e^{1n}\varepsilon^n}{1} = (e\varepsilon)^n,$$

which is the desired result. □

Exercise 2.3.2
Here we want to prove the lower-tail version of Theorem 2.3.1. We observe that for $S_N = \sum_{i=1}^{N} X_i$ we have the following:

$$S_N \leq t \Leftrightarrow -S_N \geq -t \Leftrightarrow \sum_{i=1}^{N} (-X_i) \geq -t.$$

Substituting $Y_i = -X_i$ and $t' = -t$, we see that the inequality has now the right form of a sum of Random Variables being larger than a quantity.

$$P_0 = P(S_N \leq t) = P(-S_N \geq -t) = P\left(\sum_{i=1}^{N} Y_i \geq t'\right) = P({S_N}' \geq t').$$

Starting from this, we can again follow the steps of Theorem 2.2.2: exponentiate and apply Markov's inequality and independence. This gives:

$$P_0 = P\left(\sum_{i=1}^{N} Y_i \geq t'\right) \leq e^{-\lambda t'} \prod_{i=1}^{N} \mathbb{E}\, e^{\lambda Y_i}. \qquad (2.3.2a)$$

Now, we will try to bound the MGF of the variables Y_i as in the proof of Theorem 2.3.1. We have that $Y_i = -X_i$ takes the value -1 with probability p_i, and 0 otherwise. Therefore:

$$\mathbb{E}e^{\lambda Y_i} = (1 - p_i)e^{0\lambda} + p_i e^{-1\lambda} = 1 + \left(e^{-\lambda} - 1\right)p_i \leq \exp\left[\left(e^{-\lambda} - 1\right)p_i\right],$$

where in the last step we used the inequality $e^x \geq 1 + x$ (valid for all $x \in \mathbb{R}$) with $x = \left(e^{-\lambda} - 1\right)p_i$. Substituting this in $(2.3.2a)$, we have:

$$P_0 \leq e^{-\lambda t'} \prod_{i=1}^{N} \mathbb{E}\, e^{\lambda Y_i} \leq e^{-\lambda t'} \prod_{i=1}^{N} \exp\left[\left(e^{-\lambda} - 1\right)p_i\right] = e^{-\lambda t'} \exp\left[\left(e^{-\lambda} - 1\right)\sum_{i=1}^{N} p_i\right] \Leftrightarrow$$

$$P_0 \leq e^{-\lambda t'} \exp\left[\left(e^{-\lambda} - 1\right)\mu\right],$$

since we have that $\mathbb{E}X_i = p_i \Rightarrow \mathbb{E}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} p_i = \mu$, the mean of the sum of X_i's. Putting back $t' = -t$, and setting $\lambda = \ln\left(\frac{\mu}{t}\right)$, which is positive since $t < \mu$ by assumption (hence $\frac{\mu}{t} > 1$ and $\ln\left(\frac{\mu}{t}\right) > \ln(1) = 0$), we obtain:

$$P_0 \leq e^{\lambda t} \exp\left[\left(e^{-\lambda} - 1\right)\mu\right] = \exp\left[\ln\left(\frac{\mu}{t}\right)t\right] \exp\left[\left(\frac{t}{\mu} - 1\right)\mu\right] = \left(\frac{\mu}{t}\right)^t e^{t-\mu} \Leftrightarrow$$

$$P_0 = P\left(\sum_{i=1}^{N} X_i \leq t\right) \leq e^{-\mu}\left(\frac{e\mu}{t}\right)^t,$$

as required. □

Exercise 2.3.3

Consider the Poisson random variable $X \sim \mathrm{Pois}(\lambda)$. Following the hint, we will use Theorem 1.3.4. Let $X_{N,i} \sim \mathrm{Bern}(p_{N,i})$ be Bernoulli random variables, and $S_N = \sum_{i=1}^{N} X_{N,i}$. From 1.3.4, we know that if $\max_{i \leq N} p_{N,i} \to 0$ and $\mathbb{E}S_N = \sum_{i=1}^{N} p_{N,i} \to \lambda$ as $N \to \infty$, then $S_N \to \mathrm{Pois}(\lambda)$.

Fix now some $t > \lambda$. Since $\mathbb{E}S_N \to \lambda$, then there exist some N_0 so that $\mathbb{E}S_N = {\lambda_N}' < t$ for $N \geq N_0$ (from the limit definition). For all such N, we can apply the Chernoff bound on S_N, which gives:

$$P(S_N \geq t) \leq e^{-{\lambda_N}'}\left(\frac{e{\lambda_N}'}{t}\right)^t.$$

As $N \to \infty$, ${\lambda_N}' \to \lambda$ and $S_N \to \mathrm{Pois}(\lambda)$, so S_N behaves as a Poisson variable with parameter λ in the limit. Combining these, we get:

$$P(X \geq t) = [P(X \geq t) - P(S_N \geq t)] + P(S_N \geq t) \leq f_N(t) + e^{-{\lambda_N}'}\left(\frac{e{\lambda_N}'}{t}\right)^t,$$

where $f_N(t) = P(X \geq t) - P(S_N \geq t)$. As $N \to \infty$ we have $S_N \to \mathrm{Pois}(\lambda)$ so $f_N(t) = P(X \geq t) - P(S_N \geq t) \to 0$. Taking limits on both sides, we get:

$$P(X \geq t) \leq e^{-\lambda}\left(\frac{e\lambda}{t}\right)^t,$$

as required. □

Exercise 2.3.5

Consider S_N as in the setting of theorem 2.3.1. First, we have that:

$$|S_N - \mu| \geq \delta\mu \Leftrightarrow S_N - \mu \geq \delta\mu \wedge\ S_N - \mu \leq -\delta\mu \Leftrightarrow S_N \geq (1+\delta)\mu\ \wedge S_N \leq (1-\delta)\mu,$$

where $\wedge$ is the OR operator, meaning that either of these cases can hold. From the union bound, we have:

$$P_0 = P(|S_N - \mu| \geq \delta\mu) = P(S_N \geq (1+\delta)\mu\ \wedge S_N \leq (1-\delta)\mu) \Rightarrow$$

$$P_0 \leq P(S_N \geq (1+\delta)\mu) + P(S_N \leq (1-\delta)\mu).$$

Therefore, it suffices to bound the two probabilities, $P(S_N \geq (1+\delta)\mu)$ and $P(S_N \leq (1-\delta)\mu)$. For the 1st case, we can apply the Chernoff bound with $t = (1+\delta)\mu > \mu$ (since $\delta > 0$), and we get:

$$P(S_N \geq (1+\delta)\mu) \leq e^{-\mu}\left(\frac{e\mu}{(1+\delta)\mu}\right)^{(1+\delta)\mu} = e^{-\mu}\left(\frac{e}{(1+\delta)}\right)^{(1+\delta)\mu} \qquad (2.3.5a)$$

Similarly, for the 2nd case, we can apply the version of exercise 2.3.2 of Chernoff's bound (lower tail) with $t = (1-\delta)\mu < \mu$ (since $\delta > 0$), which gives:

$$P(S_N \leq (1-\delta)\mu) \leq e^{-\mu}\left(\frac{e\mu}{(1-\delta)\mu}\right)^{(1-\delta)\mu} = e^{-\mu}\left(\frac{e}{(1-\delta)}\right)^{(1-\delta)\mu} \qquad (2.3.5b)$$

To continue now, we must show (adding both equations above):

$$e^{-\mu}\left(\frac{e}{(1+\delta)}\right)^{(1+\delta)\mu} + e^{-\mu}\left(\frac{e}{(1-\delta)}\right)^{(1-\delta)\mu} \leq 2e^{-c\mu\delta^2} \qquad (2.3.5c)$$

for $\delta \in (0,1]$ and c being an absolute constant. First, let's do some simplifications:

$$e^{-\mu}\left(\frac{e}{(1+\delta)}\right)^{(1+\delta)\mu} = e^{-\mu}e^{\mu+\mu\delta}(1+\delta)^{-(1+\delta)\mu} = e^{\mu\delta}(1+\delta)^{-(1+\delta)\mu}.$$

Similarly,

$$e^{-\mu}\left(\frac{e}{(1-\delta)}\right)^{(1-\delta)\mu} = e^{-\mu\delta}(1-\delta)^{-(1-\delta)\mu}.$$

Observing the form of the equations above, it turns out that it suffices to show:

$$e^{\mu x}(1+x)^{-(1+x)\mu} \leq e^{-c\mu x^2}, x \in [-1,1]. \qquad (2.3.5d)$$

Indeed, if that holds, then substituting δ and $-\delta$ and adding gives us $(2.3.5c)$. So, our task now is to show $(2.3.5d)$.

We have that:

$$e^{\mu x}(1+x)^{-(1+x)\mu} = e^{\mu x}\frac{1}{(1+x)^{(1+x)\mu}} \leq \frac{1}{e^{c\mu x^2}} \Leftrightarrow (1+x)^{(1+x)\mu} \geq e^{\mu x + c\mu x^2} \Leftrightarrow$$

$$(1+x)\mu \cdot \ln(1+x) \geq \mu x + c\mu x^2 \Leftrightarrow (1+x) \cdot \ln(1+x) \geq x + cx^2. \qquad (2.3.5e)$$

So, consider now the function:

$$f(x) = (1+x) \cdot \ln(1+x) - x - cx^2, x \in [-1,1].$$

We have:

$$f'(x) = \ln(1+x) + 1 - 1 - 2cx = \ln(1+x) - 2cx,$$

$$f''(x) = \frac{1}{1+x} - 2c.$$

Observe that $f'(-1) \to -\infty$, and $f'(1) = ln2 - 2c > 0$ for $c < ln2/2$. So, since $f'(x)$ is continuous, there exist a $x^* \in (-1,1)$ such that $f'(x^*) = 0$. By choosing $c < 1/4$, we can see that:

$$f''(x^*) = \frac{1}{1+x^*} - 2c > \frac{1}{1+1} - 2c = \frac{1}{2} - 2c > 0,$$

since $x^* < 1 \Leftrightarrow 1 + x^* < 1 + 1 = 2 \Leftrightarrow \frac{1}{1+x^*} > \frac{1}{2}$. Therefore, the point x^* is a local minimum. We want $f(x)$ to be positive in $(-1,1)$, hence in particular the minimum must be positive. This gives:

$$f(x^*) = (1+x^*) \cdot \ln(1+x^*) - x^* - cx^{*2} > 0 \Leftrightarrow c < \frac{(1+x^*) \cdot \ln(1+x^*) - x^*}{x^{*2}},$$

so for this c (which is a constant), we have that $f(x) > 0$ in $(-1,1)$. What remains is to check also the endpoints, e.g. it must also hold $f(-1) \geq 0, f(1) \geq 0$. We have:

$$f(1) = 2ln2 - 1 - c \geq 0 \Leftrightarrow c \leq 2ln2 - 1,$$

$$f(-1) \to 1 - c > 0 \Leftrightarrow c < 1.$$

This value for $f(-1)$ above is obtained by observing that:

$$\lim_{x \to -1} (1+x) \cdot \ln(1+x) = \lim_{u \to 0} u \cdot \ln(u) = \lim_{u \to 0} \frac{\ln(u)}{1/u} = \lim_{u \to 0} \frac{1/u}{-1/u^2} = \lim_{u \to 0} -u = 0,$$

using del' Hospital's rule. Summing up, we can combine all the inequalities in the form $c < const$ we have above and find a c so that $f(x) \geq 0$ for x in $(-1,1)$, as well as for $x = \pm 1$. Hence, there exist an absolute constant c so that $f(x) \geq 0$ for $x \in [-1,1]$, which implies that $(2.3.5e)$ holds. This in turn implies $(2.3.5d)$, which implies $(2.3.5c)$. This completes the proof. □

Exercise 2.3.6

Consider the sum of Bernoulli's S_N of Theorem 1.3.4 that approximates the Poisson distribution, and tends to it in the limit. Now, for such an S_N with mean λ', we can apply exercise 2.3.5 and we have that:

$$P_0 = P(|S_N - \lambda'| \geq \delta\lambda') \leq 2e^{-c\lambda'\delta^2} \Leftrightarrow$$

$$P(|S_N - \lambda'| \geq t') \leq 2e^{-c\lambda'\left(\frac{t\prime}{\lambda'}\right)^2} = 2\exp\left(-\frac{ct'^2}{\lambda'}\right), \tag{2.3.6a}$$

with $t' = \delta\lambda'$, $\delta \in (0,1]$. So, we have the proposition, but for the Bernoulli sum.

Now, fix a $\delta = t/\lambda$, and set $\lambda_N = \mathbb{E}S_N$, $t_N = \delta\lambda_N$. For the Poisson random variable X, write:

$$P(|X - \lambda| \geq t) = [P(|X - \lambda| \geq t) - P(|S_N - \lambda| \geq t)] + P(|S_N - \lambda| \geq t) \Rightarrow$$

$$P(|X - \lambda| \geq t) = f_N(t) + P(|S_N - \lambda| \geq t).$$

As $N \to \infty$, $f_N(t) \to 0$, since S_N converges to X in distribution.

Consider now the sets:

$$U_N = \{S_N: |S_N - \lambda| \geq t\}, U'_N = \{S_N: |S_N - \lambda_N| \geq t_N\}$$

We have that $P(U'_N) \leq 2\exp\left(-\frac{ct_N{}^2}{\lambda_N}\right)$ from above. Moreover, as $N \to \infty$, we have that $t_N \to t, \lambda_N \to \lambda$, and $P(U'_N) \to P(U_N)$. Hence, it must be the case that $P(U_N) \leq 2\exp\left(-\frac{ct^2}{\lambda}\right)$, which implies that $P(|S_N - \lambda| \geq t) \leq 2\exp\left(-\frac{ct^2}{\lambda}\right)$ as $N \to \infty$, and the conclusion follows. □

Exercise 2.3.8
Consider λ Poisson distributions $X_i \sim \mathrm{Pois}(1), i = 1, \dots, \lambda$, and consider $S_\lambda = \sum_{i=1}^{\lambda} X_i$. We have that $\mathbb{E}S_\lambda = \sum_{i=1}^{\lambda} \mathbb{E}X_i = \sum_{i=1}^{\lambda} 1 = \lambda$, $Var[S_\lambda] = \sum_{i=1}^{\lambda} Var[X_i] = \sum_{i=1}^{\lambda} 1 = \lambda$, since a Poisson with parameter k has mean k and variance k.

With these, the central limit theorem tells us that

$$\frac{S_\lambda - \mathbb{E}S_\lambda}{\sqrt{Var[S_\lambda]}} = \frac{S_\lambda - \lambda}{\sqrt{\lambda}} \to N(0,1)$$

in distribution, as $\lambda \to \infty$. But we know that the sum of independent Poisson random variables is again a Poisson, so we have that $S_\lambda \sim \mathrm{Pois}(\lambda)$. From this, the conclusion follows. □

Exercise 2.4.2

We are given a random graph $G \sim G(n,p)$ with expected degree d. We are given that $d = O(\ln n)$, that is, there is a given constant C such that $d \leq C \ln n$ for sufficiently large n. Let d_{max} be the maximum degree of G. We want to show that with high probability, e.g. 0.90, we have that $d_{max} = O(\ln n)$, e.g. there is a constant C' so that $d_{max} \leq C' \ln n$, with probability 0.9.

The complementary event here is the following: for every $c \in \mathbb{R}^+$, there is a sufficiently large n so that $d_{max} \geq c \ln n$. Clearly, d_{max} in this case must be larger than d (otherwise, d_{max} is also $O(\ln n)$ and we are done). Since the degree of a node u is a sum of $\mathrm{Ber}(p)$ random variables with expected value d, we can apply the Chernoff bound with $t = d_{max} > d, \mu = d$, and we get:

$$P_0 = P(d_u \geq d_{max}) \leq e^{-d} \left(\frac{ed}{d_{max}}\right)^{d_{max}}.$$

For sufficiently large n's, we have $d_{max} \geq c \ln n$ for every c, so we can choose $c = kCe$ for a $k > 0$, and we find:

$$P_0 \leq e^{-d} \left(\frac{ed}{d_{max}}\right)^{d_{max}} \leq \left(\frac{ed}{d_{max}}\right)^{d_{max}} \quad | \text{ since } e^{-d} \leq 1, \text{bec.}\, d \geq 0$$

$$\Rightarrow P_0 \leq \left(\frac{eC \ln n}{d_{max}}\right)^{d_{max}} \quad | \text{ since } d \leq C \ln n$$

$$\Rightarrow P_0 \leq \left(\frac{eC \ln n}{kCe \ln n}\right)^{d_{max}} = \left(\frac{1}{k}\right)^{d_{max}} = k^{-d_{max}} \,| \text{ since } d_{max} \geq kCe \ln n$$

$$\Rightarrow P_0 \leq k^{-d_{max}} \leq k^{-kCe \ln n} \,| \text{ since } d_{max} \geq kCe \ln n \text{ and } 2^{-x} \text{ is decreasing}$$

Now, we choose a k large enough so that $kCe \geq 2$ and $k \geq e$ (e.g. choose $k = \max\left(e, \frac{2}{Ce}\right)$). With that k, we have:

$$P_0 \leq k^{-kCe \ln n} \leq e^{-kCe \ln n} \,| \text{ since } k \geq e \text{ and } x^{-a} \text{ is decreasing for } a > 0$$

$$\Rightarrow P_0 \leq e^{-kCe \ln n} \leq e^{-2 \ln n} = e^{-\ln n^2} = \frac{1}{n^2}.$$

But $\frac{1}{n^2} \leq \frac{1}{10n}$ for sufficiently large n. So, we finally get:

$$P_0 = P(d_u \geq d_{max}) \leq \frac{1}{10n}.$$

Finally, using the union bound, we get:

$$P(\exists u\colon d_u \geq d_{max}) \leq \sum_{j=1}^{n} P\big(d_j \geq d_{max}\big) \leq n\frac{1}{10n} = 0.1.$$

Hence, the complementary event, $d_{max} = O(\ln n)$, happens with probability at least 90%. This is what we wanted to show. □

Exercise 2.4.3

We are given a random graph $G \sim G(n, p)$ with expected degree d. We know that $d = O(1)$, e.g. there is a (given) constant C such that $d \leq C$ for sufficiently large n. Let d_{max} be the maximum degree of G. We want to show that with high probability, e.g. 0.90, we have that $d_{max} = O(\frac{\ln n}{\ln \ln n})$, e.g. there is a constant C' so that $d_{max} \leq C' \frac{\ln n}{\ln \ln n}$, with probability at least 0.9.

The complementary event here is the following: for every $c \in \mathbb{R}^+$, there is a sufficiently large n so that $d_{max} \geq c \frac{\ln n}{\ln \ln n}$. Clearly, d_{max} in this case must be larger than d (otherwise, d_{max} is also $O(1)$ and we are done). Since the degree of a node u is a sum of $\mathrm{Ber}(p)$ random variables with expected value d, we can apply the Chernoff bound with $t = d_{max} > d, \mu = d$, and we get:

$$P_0 = P(d_u \geq d_{max}) \leq e^{-d}\left(\frac{ed}{d_{max}}\right)^{d_{max}}.$$

For sufficiently large n's, we have $d_{max} \geq c \frac{\ln n}{\ln \ln n}$ for every c, so we can choose $c = kCe$ for a $k > 0$, and we find:

$$P_0 \leq e^{-d}\left(\frac{ed}{d_{max}}\right)^{d_{max}} \leq \left(\frac{ed}{d_{max}}\right)^{d_{max}} \quad |\ \text{since } e^{-d} \leq 1, \text{bec. } d \geq 0$$

$$\Rightarrow P_0 \leq \left(\frac{eC}{d_{max}}\right)^{d_{max}} \quad |\ \text{since } d \leq C$$

$$\Rightarrow P_0 \leq \left(\frac{eC}{kCe\frac{\ln n}{\ln \ln n}}\right)^{d_{max}} = \left(\frac{\ln \ln n}{k \ln n}\right)^{d_{max}} \quad |\ \text{since } d_{max} \geq kCe\frac{\ln n}{\ln \ln n}$$

We can choose k so that $kCe = k' \geq 1$. With this choice, we have:

$$P_0 \leq \left(\frac{\ln \ln n}{k \ln n}\right)^{k'\frac{\ln n}{\ln \ln n}} = \left(\frac{k \ln n}{\ln \ln n}\right)^{-k'\frac{\ln n}{\ln \ln n}} \quad |\ \text{since } d_{max} \geq kCe\frac{\ln n}{\ln \ln n} \geq k'\frac{\ln n}{\ln \ln n}, \text{bec. } kCe \geq k'$$

Now, we want to bound this by $\frac{1}{10n}$. To do this, substitute $\frac{\ln n}{\ln \ln n} = x$, and solve:

$$(kx)^{-k'x} \leq \frac{1}{10n} \Leftrightarrow -k'x \ln kx \leq -\ln 10n = \text{ | taking logs}$$

$$\Leftrightarrow k'x \ln kx \geq \ln(10n)$$

Plugging back $x = \frac{\ln n}{\ln \ln n}$, we find:

$$A = k' \frac{\ln n}{\ln \ln n} \ln k \left(\frac{\ln n}{\ln \ln n}\right) \geq \ln(10n) \qquad (2.4.3a)$$

Simplifying, we get:

$$A = k' \frac{\ln n}{\ln \ln n} \ln k \left(\frac{\ln n}{\ln \ln n}\right) = k' \frac{\ln n}{\ln \ln n} [\ln k + \ln \ln n - \ln \ln \ln n] \geq \ln 10n \text{ | ln properties}$$

$$\Leftrightarrow \frac{k' \ln k}{\ln \ln n} + k' - \frac{k' \ln \ln \ln n}{\ln \ln n} \geq 1 + \frac{\ln 10}{\ln n} \text{ | dividing by } \ln n \text{ and ln properties}$$

$$\Leftrightarrow \frac{k' \ln k}{\ln \ln n} + (k' - 1) - \frac{k' \ln \ln \ln n}{\ln \ln n} \geq \frac{\ln 10}{\ln n} \Leftrightarrow (k' - 1) \geq \frac{\ln 10}{\ln n} - \frac{k'^{\ln k}}{\ln \ln n} + \frac{k' \ln \ln \ln n}{\ln \ln n}$$

This holds for sufficiently large n, since $\frac{\ln \ln \ln n}{\ln \ln n} \to 0, \frac{\ln 10}{\ln n} \to 0, \frac{k' \ln k}{\ln \ln n} \to 0$ as $n \to \infty$, while $k' - 1 > 0$ choosing $k' > 1$. Going back to P_0, we established that:

$$P_0 = P(d_u \geq d_{max}) \leq \frac{1}{10n} \text{ for large } n.$$

Finally, using the union bound, we get:

$$P(\exists u \colon d_u \geq d_{max}) \leq \sum_{j=1}^{n} P\left(d_j \geq d_{max}\right) \leq n \frac{1}{10n} = 0.1.$$

Hence, the complementary event, $d_{max} = O\left(\frac{\ln n}{\ln \ln n}\right)$, happens with probability at least 90%. This is what we wanted to show. □

Exercise 2.4.4

Each degree d_i is a Binomial R.V., e.g. $p_i = Bin(n-1, p)$, but the d_i's are not independent, since if one vertex u connects to one vertex v, then u affects the degree of v.

To fix this, we will take a subset of the vertices at random, such as no two vertices are connected to each other (with high probability). Specifically, we select $n^{1/3}$ vertices at random, and form the set $V' \subset V$.

Since for the expected degree we know that $d = o(\log n) = (n-1)p$, we find that $p = \frac{o(\log n)}{n-1} < \frac{o(\log n)}{\frac{1}{2}n} = o\left(\frac{\log n}{n}\right)$, e.g. the probability that two vertices are connected is bounded by $o\left(\frac{\log n}{n}\right)$. Thus, using the union bound, the probability that any two nodes $u, v \in V'$ are connected, satisfies:

$$\mathbb{P}[\exists u, v \in V' : (u, v) \in G] \leq p \cdot |V'|^2 = o\left(\frac{\log n}{n}\right) \cdot \left(n^{\frac{1}{3}}\right)^2 = o\left(\frac{\log n}{n^{\frac{1}{3}}}\right) \equiv p_1(n).$$

Since $\frac{\log n}{n^{\frac{1}{3}}} \to 0$ for $n \to \infty$, the probability that some nodes in V' are connected can become as small as we like. In the equation above, (u, v) devotes an edge, and we take the probability that this edge belongs to the graph G.

So, assume now that we have a subset V' of unconnected nodes. Within V', the degrees d_i of the vertices are now independent. Therefore, for the probability that a node has degree $10d$, we find:

$$P = \mathbb{P}[\exists i \in V' : d_i = 10d] = 1 - \mathbb{P}[\forall i \in V' : d_i \neq 10d] = 1 - (\mathbb{P}[d_i \neq 10d])^{|V'|} \Rightarrow$$

$$P = \mathbb{P}[\exists i \in V' : d_i = 10d] \geq 1 - (\mathbb{P}[d_i \neq 10d])^n \qquad (2.4.4.1)$$

where we used that $\mathbb{P}[\forall i \in V' : d_i \neq 10d] = (\mathbb{P}[d_i \neq 10d])^{|V'|}$ due to the independence of the d_i's, and further that $|V'| \leq n \Rightarrow (\mathbb{P}[d_i \neq 10d])^{|V'|} \geq (\mathbb{P}[d_i \neq 10d])^n$.

Now, since d_i is a sum of Bernoulli R.V.s with small probability, we can use the Poisson approximation for d_i, noting that for the Poisson parameter λ we have $\lambda = \mathbb{E}[d_i] = d$:

$$\mathbb{P}[d_i = 10d] = \frac{1}{\sqrt{2\pi 10d}} e^{-d} \left(\frac{ed}{10d}\right)^{10d} + o(1) \equiv q(n) + o(1)$$

Here we assumed that $10d$ is an integer (the Poisson distribution is defined only for integer values), and further $o(1)$ is a correction factor for the Poisson approximation, that goes to zero as $n \to \infty$.

Working on the above formula, and remembering that $d = o(\log n) \Leftrightarrow d \leq c \log n$ for any $c > 0$ for sufficiently large n, we get:

$$q(n) = \frac{1}{\sqrt{2\pi 10d}} e^{-d} \left(\frac{ed}{10d}\right)^{10d} = \frac{1}{\sqrt{2\pi 10d}} e^{-d} \left(\frac{e}{10}\right)^{10d} \Leftrightarrow$$

$$q(n) = e^{-\frac{1}{2}\log(2\pi 10d)} e^{-d} e^{10d \log\left(\frac{e}{10}\right)} \Leftrightarrow$$

$$q(n) = \exp\left(10d\, log\left(\frac{e}{10}\right) - d - \frac{1}{2} log(2\pi 10d)\right) \geq \exp\left(-d - \frac{1}{2} log(2\pi 10d)\right) \Rightarrow$$

$$q(n) \geq \exp\left(-c \log n - \frac{1}{2} log(2\pi 10c \log n)\right) \geq \exp(-2c \log n) = n^{-2c} = n^{-c'}$$

since for sufficiently large n we have $c \log n \geq \frac{1}{2} log(2\pi 10c \log n)$.

Thus we have:

$$\mathbb{P}[d_i = 10d] = q(n) + o(1) \geq n^{-c'} + o(1),$$

And plugging this back into (2.4.4.1) we find:

$$P \geq 1 - (\mathbb{P}[d_i \neq 10d])^n \geq 1 - (1 - \mathbb{P}[d_i = 10d])^n \geq 1 - \left(1 - n^{-c'} - o(1)\right)^n \Rightarrow$$

$$P \geq 1 - \left(e^{-n^{-c'}} - o(1)\right)^n.$$

The right hand side will become larger than say 0.9, for a sufficiently large n. In the last step we used the inequality $e^x \geq 1 + x$ (which is valid for all $x \in \mathbb{R}$) with $x = n^{-c'}$.

To be more precise, consider the complementary probability $P' = 1 - P \leq \left(e^{-n^{-c'}} - o(1)\right)^n$. Also, from the previous considerations, remember that we will fail to pick an independent set V' with probability at most $p_1(n) = o\left(\frac{\log n}{n^{\frac{1}{3}}}\right)$. By the union bound, the overall failure probability will satisfy:

$$P_{fail} \leq P' + p_1(n) \leq \left(e^{-n^{-c'}} - o(1)\right)^n + o\left(\frac{\log n}{n^{\frac{1}{3}}}\right).$$

For sufficiently large n we see that $P_{fail} \leq 0.1$, which concludes the proof (e.g. we will find a vertex satisfying the requirement with probability at last 0.9).

Exercise 2.4.5

We 'll use a similar approach as in exercise 2.4.4. Again, since we know that the expected degree is $d = O(1)$ we find: $d = (n-1)p \leq C \Leftrightarrow p \leq \frac{C}{n-1} \Rightarrow p = O\left(\frac{1}{n}\right)$. Therefore, by picking a random subset V' of vertices of size $|V'| = n^{1/3}$, we have that the probability that two nodes in V' are connected is:

$$\mathbb{P}[\exists u, v \in V': (u,v) \in G] \leq p \cdot |V'|^2 = O\left(\frac{1}{n}\right) \cdot \left(n^{\frac{1}{3}}\right)^2 = O\left(\frac{1}{n^{\frac{1}{3}}}\right) \equiv p_1(n),$$

which tends to 0 for large n. Therefore, we may assume that the vertices in V' are unconnected, and their degrees d_i independent.

By independence, for the probability that a node has degree $d' = c\frac{\log n}{\log\log n}$, we find:

$$P = \mathbb{P}[\exists i \in V': d_i = d'] = 1 - \mathbb{P}[\forall i \in V': d_i \neq d'] = 1 - (\mathbb{P}[d_i \neq d'])^{|V'|} \Rightarrow$$

$$P = \mathbb{P}[\exists i \in V': d_i = 10d] \geq 1 - (\mathbb{P}[d_i \neq d'])^n \tag{2.4.5.1}$$

where we used that $\mathbb{P}[\forall i \in V': d_i \neq d'] = (\mathbb{P}[d_i \neq d'])^{|V'|}$ due to the independence of the d_i's, and further that $|V'| \leq n \Rightarrow (\mathbb{P}[d_i \neq d'])^{|V'|} \geq (\mathbb{P}[d_i \neq d'])^n$.

Now, since d_i is a sum of Bernoulli R.V.s with small probability, we can use the Poisson approximation for d_i, noting that for the Poisson parameter λ we have $\lambda = \mathbb{E}[d_i] = d$:

$$\mathbb{P}[d_i = d'] = \frac{1}{\sqrt{2\pi d'}} e^{-d} \left(\frac{ed}{d'}\right)^{d'} + o(1) \equiv q(n) + o(1)$$

Here we assumed that d' is an integer (the Poisson distribution is defined only for integer values), and further $o(1)$ is a correction factor for the Poisson approximation, that goes to zero as $n \to \infty$.

Now, for $q(n)$ we find:

$$q(n) = \frac{1}{\sqrt{2\pi d'}} e^{-d} \left(\frac{ed}{d'}\right)^{d'} = \exp\left(-d - \frac{1}{2} log\, 2\pi d' + d'\, log\, \frac{ed}{d'}\right),$$

and plugging $d \leq C, d' = c\frac{\log n}{\log\log n}$ this becomes:

$$q(n) = \exp\left(-d - \frac{1}{2} log\, 2\pi d' + d'\, log\, \frac{ed}{d'}\right) \Rightarrow$$

$$q(n) = \exp\left(-d - \frac{1}{2} log\, 2\pi d' + d'\, log\, ed - d' \log d'\right) \Rightarrow$$

$$q(n) \geq \exp\left(-\frac{1}{2} d' \log d'\right),$$

since for sufficiently large n we have that $-d-\frac{1}{2} log\, 2\pi d' + d'\, log\, ed \le \frac{1}{2} d' \log d'$ (because $d \le C$ and $d' = c\frac{\log n}{\log\log n}$). This gives:

$$q(n) \ge \exp\left(-\frac{1}{2} d' \log d'\right) \ge \exp\left(-\frac{1}{2} c\frac{\log n}{\log\log n}\log\left(c\frac{\log n}{\log\log n}\right)\right) \Rightarrow$$

$$q(n) \ge \exp\left(-c'\frac{\log n}{\log\log n}(\log c + \log\log n - \log\log\log n)\right) \Rightarrow$$

$$q(n) \ge \exp\left(-c'\frac{\log n \log c}{\log\log n} - c' \log n + c'\frac{\log n \log\log\log n}{\log\log n}\right),$$

and for sufficiently large n the term $c' \log n$ will dominate (above, we set $\frac{1}{2} c = c'$). Therefore, for some other constant c'' we will have:

$$q(n) \ge \exp\left(-c''^{\log n}\right) \ge n^{-c''}.$$

Thus we have:

$$\mathbb{P}[d_i = d'] = q(n) + o(1) \ge n^{-c''} + o(1),$$

And plugging this back into (2.4.5.1) we find:

$$P \ge 1 - (\mathbb{P}[d_i \ne d'])^n \ge 1 - (1 - \mathbb{P}[d_i = d'])^n \ge 1 - \left(1 - n^{-c''} - o(1)\right)^n \Rightarrow$$

$$P \ge 1 - \left(e^{-n^{-c''}} - o(1)\right)^n.$$

The right hand side will become larger than say 0.9, for a sufficiently large n. In the last step we used the inequality $e^x \ge 1 + x$ (which is valid for all $x \in \mathbb{R}$) with $x = n^{-c''}$.

To be more precise, consider the complementary probability $P' = 1 - P \le \left(e^{-n^{-c''}} - o(1)\right)^n$. Also, from the previous considerations, remember that we will fail to pick an independent set V' with probability at most $p_1(n) = O\left(\frac{1}{n}\right)$. By the union bound, the overall failure probability will satisfy:

$$P_{fail} \le P' + p_1(n) \le \left(e^{-n^{-c''}} - o(1)\right)^n + O\left(\frac{1}{n}\right).$$

For sufficiently large n we see that $P_{fail} \le 0.1$, which concludes the proof (e.g. we will find a vertex satisfying the requirement with probability at last 0.9).