STATS 214 Autumn 2021 Homework 1

SUNet ID: 06009508

Name: Brandon McKinzie

Collaborators: N/A

By turning in this assignment, I agree by the Stanford honor code and declare that all of this is my own work.

Problem 1 (a) Sharp bounds on sub-Gaussianity of bounded variables

Suppose a random variable X is between $[a, b]$ almost surely. Prove that X is a sub-Gaussian random variable with variance proxy $\frac{(a-b)^2}{4}$.

1. Let $\psi(\lambda) = \log \mathbb{E}\left[e^{\lambda X}\right]$. We'll evaluate $\psi(0)$ its derivative $\psi'(0)$.

$$\psi(0) = \log \mathbb{E}\left[e^{0}\right] = \log 1 = 0 \tag{1}$$

$$\psi'(0) = \frac{\partial}{\partial\lambda} \log \mathbb{E}\left[e^{\lambda X}\right]\Big|_{\lambda=0} \tag{2}$$

$$= \frac{1}{\mathbb{E}\left[e^{\lambda X}\right]} \int \mathrm{d}x \frac{\partial}{\partial\lambda} p(X{=}x) e^{\lambda x}\Big|_{\lambda=0} \tag{3}$$

$$= \frac{1}{\mathbb{E}\left[e^{\lambda X}\right]} \int \mathrm{d}x p(X{=}x) x e^{\lambda x}\Big|_{\lambda=0} \tag{4}$$

$$= \frac{1}{\mathbb{E}\left[e^{0}\right]} \int \mathrm{d}x p(X{=}x) x e^{0} \tag{5}$$

$$= \mathbb{E}\left[X\right] \tag{6}$$

2. Let $\mathbb{E}_\lambda\left[f(X)\right] := \mathbb{E}\left[f(X)e^{\lambda X}\right] / \mathbb{E}\left[e^{\lambda X}\right]$. First, we will show that $\psi''(\lambda) = \mathbb{E}_\lambda\left[X^2\right] - \mathbb{E}_\lambda\left[X\right]^2$.

$$\psi''(\lambda) = \frac{\partial}{\partial\lambda}\left[\frac{1}{\mathbb{E}\left[e^{\lambda X}\right]} \int \mathrm{d}x p(x) x e^{\lambda x}\right] \tag{7}$$

$$= \frac{\partial}{\partial\lambda}\left[\frac{1}{\mathbb{E}\left[e^{\lambda X}\right]} \mathbb{E}\left[X e^{\lambda X}\right]\right] \tag{8}$$

$$= \frac{\mathbb{E}\left[e^{\lambda X}\right]\mathbb{E}\left[X^2 e^{\lambda X}\right] - \mathbb{E}\left[X e^{\lambda X}\right]^2}{\mathbb{E}\left[e^{\lambda X}\right]^2} \tag{9}$$

$$= \frac{\mathbb{E}\left[X^2 e^{\lambda X}\right]}{\mathbb{E}\left[e^{\lambda X}\right]} - \frac{\mathbb{E}\left[X e^{\lambda X}\right]^2}{\mathbb{E}\left[e^{\lambda X}\right]^2} \tag{10}$$

$$= \mathbb{E}_\lambda\left[X^2\right] - \mathbb{E}_\lambda\left[X\right]^2 \tag{11}$$

Next, we'll derive a bound for $|\psi''(\lambda)|$ by exploiting some basic facts about $\mathrm{Var}\left[X\right]$ for any random variable X, along with some basic inequalities exploiting the fact that X in this problem is bounded between $[a, b]$ almost surely. Our goal is to show that

$$\sup_{\lambda \in \mathbb{R}} |\psi''(\lambda)| \leq \frac{(a-b)^2}{4} \tag{12}$$

First, note that $\psi''(\lambda)$ looks eerily like a variance of some kind. Informally, it is the variance of X if the probability distribution $p(x)$ was scaled by $\exp \lambda x$ at each point, then normalized by dividing $\mathbb{E}\left[e^{\lambda x}\right]$. We can show that, for a bounded random variable X that takes on values within $[a, b]$, the value of c that minimizes $\mathbb{E}\left[(X-c)^2\right]$ is $\mathbb{E}[X]$:

$$\begin{aligned} \frac{d}{dc}\mathbb{E}\left[(X-c)^2\right] &= -2\mathbb{E}[X-c] = 0 \quad \rightarrow \quad c = \mathbb{E}[X] && (13) \\ \mathrm{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] && (14) \\ &\leq \mathbb{E}\left[(X-c)^2\right] \qquad \text{for } (a \leq c \leq b) && (15) \end{aligned}$$

If we take $c = (a+b)/2$, then this yields

$$\begin{aligned} \mathrm{Var}(X) &\leq \mathbb{E}\left[\left(\frac{1}{2}(2X-a-b)\right)^2\right] && (16) \\ &= \frac{1}{4}\mathbb{E}\left[((X-a)+(X-b))^2\right] && (17) \\ &\leq \frac{1}{4}\mathbb{E}\left[((X-a)-(X-b))^2\right] \quad \text{since } X \in [a,b] \text{ a.s.} && (18) \\ &= \frac{1}{4}\mathbb{E}\left[(b-a)^2\right] && (19) \\ &= \frac{(b-a)^2}{4} && (20) \end{aligned}$$

Note that this did not rely on a particular value of λ. The value of λ only influences the weights in the aforementioned re-weighted probability distribution. Therefore, this holds for all $\lambda \in \mathbb{R}$ and we get the desired result.

3. Finally, we want to show that

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq e^{\frac{1}{2}\sigma^2\lambda^2} \quad (\forall \lambda \in \mathbb{R}) \tag{21}$$

with $\sigma^2 = (a-b)^2/4$. So far, we currently have a bound on the variance of X under the reweighted distribution parameterized by λ:

$$\psi''(\lambda) \leq \frac{(a-b)^2}{4} \tag{22}$$

and need to massage this to look more like 21. Our approach will be to twice integrate both sides and exponentiate, which will reveal the desired result. Notice that

$$
\begin{aligned}
\int_0^\lambda d\lambda' \psi''(\lambda') &= \psi'(\lambda) - \psi'(0) && (23)\\
&= \psi'(\lambda) - \mu && (24)\\
\int_0^\lambda d\lambda' (\psi'(\lambda') - \mu) &= (\psi(\lambda) - \mu\lambda) - (\psi(0) - \mu \cdot 0) && (25)\\
&= \psi(\lambda) - \lambda\mu && (26)
\end{aligned}
$$

In other words, we can twice integrate the LHS of 22 and exponentiate to get

$$
\exp\left(\psi(\lambda) - \lambda\mu\right) = \mathbb{E}\left[\exp\left(\lambda\left(X - \mu\right)\right)\right] \tag{27}
$$

All that remains is to apply the same (twice integration and exponentiation) to the RHS of 22 to obtain the desired result:

$$
\begin{aligned}
\mathbb{E}\left[e^{\lambda(X-\mu)}\right] &\leq \exp\left(\int_0^\lambda d^2\lambda \frac{(b-a)^2}{4}\right) && (28)\\
&= \exp\left(\lambda^2 \frac{(b-a)^2}{8}\right) && (29)
\end{aligned}
$$

and thus X is sub-Gaussian with variance proxy $(a-b)^2/4$.

Problem 1 (b) Sub-exponential random variables

Show the following is true for sub-exponential random variable X:

$$Pr\left[X \geq \mu + t\right] \leq \begin{cases} e^{-\frac{t^2}{2\nu^2}} & 0 \leq t < \nu^2/b \\ e^{-\frac{t}{2b}} & t \geq \nu^2/b \end{cases} \tag{30}$$

We can proceed largely the same as we did for the analogous proof with sub-Gaussian variables. First, we apply Markov's inequality on the tail bound, then utilize the definition of sub-exponential random variables. The main difference is that, when finding the optimal value of λ, we need to obey the constraint that $|\lambda| < \frac{1}{b}$. NB: in all equations below, $|\lambda| < \frac{1}{b}$.

$$\begin{aligned} \Pr\left[X - \mu \geq t\right] &= \Pr\left[e^{\lambda(X-\mu)} \geq e^{\lambda t}\right] && (31) \\ &\leq e^{-\lambda t}\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \quad \textbf{[Markov's ineq]} && (32) \\ &\leq e^{\nu^2\lambda^2/2-\lambda t} && (33) \end{aligned}$$

In the next step, we just need to be explicit when finding the minimum with respect to λ, that we must satisfy the constraint that $|\lambda| < \frac{1}{b}$:

$$\begin{aligned} \frac{d}{d\lambda}\left(\frac{1}{2}\nu^2\lambda^2 - \lambda t\right) &= \nu^2\lambda - t = 0 && (34) \\ \lambda &= \max\left(-\frac{1}{b}, \min\left(\frac{1}{b}, \frac{t}{\nu^2}\right)\right) && (35) \end{aligned}$$

We can see that when $|t/\nu^2| < \frac{1}{b}$, this leads to the tail bound for sub-Gaussian random variables, which gives us the top half of the desired result (the case where $0 \leq t < \nu^2/b$). However, for $t \geq \nu^2/b$, i.e. for $t = C\nu^2/b$ for $C \geq 1$, this yields

$$\begin{aligned} \lambda &= \max\left(-\frac{1}{b}, \min\left(\frac{1}{b}, \frac{C}{b}\right)\right) = \frac{1}{b} && (36) \\ e^{\nu^2\lambda^2/2-\lambda t} &= e^{\frac{1}{2}\frac{\nu^2}{b^2}-\frac{t}{b}} && (37) \\ &\leq e^{\frac{1}{2}\frac{t}{b}-\frac{t}{b}} && (38) \\ &= e^{-\frac{t}{2b}} && (39) \end{aligned}$$

and we have the desired result.

Problem 1 (c) Case Study

Suppose Z is a mean-zero random variable between $[-1,1]$ and var$(Z) = \gamma^2 \ll 1$.

1. From part (a), we know that, since Z is bounded, Z is sub-Gaussian. Substituting $a = -1$ and $b = 1$, we have the desired result:

$$\sigma^2 = \frac{((-1) - (1))^2}{4} = 1 \tag{40}$$

2. First, as we showed in part (a), we can use the fact that

$$\mathbb{E}_\lambda\left[(Z - \mathbb{E}_\lambda[Z])^2\right] \leq \mathbb{E}_\lambda\left[Z^2\right] \tag{41}$$

and focus on upper-bounding $\mathbb{E}_\lambda[Z^2]$. Utilizing the provided hint regarding $F_Z(u) = \Pr[Z \leq u]$, we can write

$$\mathbb{E}_\lambda\left[Z^2\right] = \int_{-1}^{1} u^2 \frac{e^{\lambda u}}{\mathbb{E}[e^{\lambda Z}]} \mathrm{d}F_Z(u) \tag{42}$$

Since the integral is over $-1 \leq u \leq 1$, we can assert the following about the fraction in the integrand (noting also that $|\lambda| \leq 1$):

$$\mathbb{E}\left[e^{\lambda Z}\right] = \int_{-1}^{1} e^{\lambda u} dF_Z(u) \geq e^{-1} \int_{-1}^{1} dF_Z(u) = e^{-1} \tag{43}$$

$$\frac{e^{\lambda u}}{\mathbb{E}[e^{\lambda Z}]} \leq \frac{e^1}{e^{-1}} = e^2 \tag{44}$$

Therefore, we can upper-bound $\mathbb{E}_\lambda[Z^2]$ as

$$\mathbb{E}_\lambda\left[Z^2\right] \leq e^2 \int_{-1}^{1} u^2 \mathrm{d}F_Z(u) = e^2\gamma^2 = O(\gamma^2) \tag{45}$$

We can then apply the same logic of problem 1(a) part (3) to get a bound on $\mathbb{E}\left[e^{\lambda Z}\right]$:

$$\mathbb{E}\left[e^{\lambda Z}\right] \leq \exp\left(\int_0^\lambda \mathrm{d}^2\lambda' e^2\gamma^2\right) = \exp\left(\frac{1}{2}\lambda^2 e^2\gamma^2\right) \quad (\forall|\lambda| \leq 1) \tag{46}$$

Therefore, Z is sub-exponential with parameter $(O(\gamma), 1)$.

3. The sub-Gaussian and sub-exponential tail bounds for Z are as follows:

$$\textbf{[sub-Gaussian]} \qquad \Pr[Z \geq t] \;\leq\; e^{-\frac{t^2}{2}} \quad (\forall t \in \mathbb{R}) \tag{47}$$

$$\textbf{[sub-exponential]} \qquad \Pr[Z \geq t] \;\leq\; \begin{cases} e^{-\frac{t^2}{2O(\gamma^2)}} & 0 \leq t < O(\gamma^2) \\ e^{-\frac{t}{2}} & t > O(\gamma^2) \end{cases} \tag{48}$$

Since $\gamma^2 << 1$, the sub-exponential bound only decays like $O(e^{-t^2})$ for a very small range of t, whereas the sub-Gaussian bound applies for all $t \in \mathbb{R}$. Also, since the tighter bound for the sub-exponential only holds for $0 \leq t \leq O(\gamma^2)$, we won't get better than $e^{-O(\gamma^2)}$ in that regime. Therefore, the sub-Gaussian bound is significantly stronger than the sub-exponential bound.

Problem 1 (d) Sub-exponential concentration

We can plug in the form of X^* into the MGF

$$\mathbb{E}\left[e^{\lambda(X^*-\mathbb{E}[X^*])}\right] = \mathbb{E}\left[e^{\lambda\left(\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_k\right)}\right] \tag{49}$$

$$= \mathbb{E}\left[e^{\left(\sum_{k=1}^n \lambda(X_k-\mu_k)\right)}\right] \tag{50}$$

$$= \prod_{k=1}^n \mathbb{E}\left[e^{\lambda(X_k-\mu_k)}\right] \qquad \textbf{[by independence]} \tag{51}$$

$$\leq \prod_{k=1}^n e^{\nu_k^2\lambda^2/2} \qquad (\forall|\lambda| < \min_k \frac{1}{b_k}) \tag{52}$$

So, assuming all $b_k > 0$[1] we have that $b^* = \max_k b_k$. Finally, this gives us the bound

$$\prod_{k=1}^n e^{\nu_k^2\lambda^2/2} = e^{\frac{\lambda^2}{2}\sum_{k=1}^n \nu_k^2} \qquad (\forall|\lambda| < \frac{1}{b^*}) \tag{53}$$

and thus X^* is sub-exponential with $\nu^* = \sqrt{\sum_{k=1}^n \nu_k^2}$. and $b^* = \max_k b_k$.

[1]Which appears to be true from other definitions, e.g. see http://www.stat.cmu.edu/ arinaldo/Teaching/36709/S19/Scribed_Lectures/Feb5_Aleksandr.pdf

Problem 1 (e) Bernstein inequality variant

Prove the following tail bound by applying part (b)

$$Pr\left[\frac{1}{n}\left(X^* - \mu^*\right) \geq t\right] \;\; \leq \;\; \begin{cases} e^{-\frac{n^2t^2}{2(\nu^*)^2}} & 0 \leq t < \frac{(\nu^*)^2}{nb^*} \\ e^{-\frac{nt}{2b^*}} & t \geq \frac{(\nu^*)^2}{nb^*} \end{cases} \tag{54}$$

If we simply plug in the result from part (b) for sub-exponential variables for the case of X^*, which we've already shown is sub-exponential, we get

$$\Pr\left[X^* \geq \mu^* + t\right] \leq \begin{cases} e^{-\frac{t^2}{2(\nu^*)^2}} & 0 \leq t < (\nu^*)^2/b^* \\ e^{-\frac{t}{2b^*}} & t \geq (\nu^*)^2/(b^*) \end{cases} \tag{55}$$

Therefore, the problem is essentially asking us to show how scaling a sub-exponential random variable by $\frac{1}{n}$ affects its tail bound. We can apply some simple arithmetic operations to see that the problem is asking us to find the tail bound for

$$\Pr\left[X^* \geq nt + \mu^*\right] \tag{56}$$

In other words, we can just replace t from part (b) with nt, immediately yielding the desired result:

$$\Pr\left[\frac{1}{n}\left(X^* - \mu^*\right) \geq t\right] \;\; \leq \;\; \begin{cases} e^{-\frac{n^2t^2}{2(\nu^*)^2}} & 0 \leq t < \frac{(\nu^*)^2}{nb^*} \\ e^{-\frac{nt}{2b^*}} & t \geq \frac{(\nu^*)^2}{nb^*} \end{cases} \tag{57}$$

Problem 1 (f) Case study II

Consider the cases where $X_1, \ldots, X_n$ are all independent distributed as the random variable Z in part (c). Derive the tail bound for X^* in two ways.

First, notice that since $\mathbb{E}[Z] = 0$, we have $\mathbb{E}[X^*] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \mu^* = 0$.

1. *Only using the fact that X_i's are sub-Gaussian with variance proxy 1.* We know that the sum of independent sub-Gaussian random variables is itself sub-Gaussian with, for this case,

$$\Pr\left[\frac{1}{n}X^* \geq t\right] \leq \exp\left(-\frac{n^2t^2}{2\sum_{i=1}^{n}\sigma_i^2}\right) = \exp\left(-\frac{nt^2}{2}\right) \quad (\forall t \in \mathbb{R}) \tag{58}$$

2. *Using the fact that X_i are sub-exponential with parameter (ν_i, b_i) where $\nu_i = O(\gamma)$, $b_i = 1$.* In this case, we know that the sum X^* is also sub-exponential with parameters (ν^*, b^*) of the form derived in part (d), with associated tail bound

$$\Pr\left[\frac{1}{n}X^* \geq t\right] \leq \begin{cases} e^{-\frac{n^2t^2}{2(\nu^*)^2}} & 0 \leq t < \frac{(\nu^*)^2}{nb^*} \\ e^{-\frac{nt}{2b^*}} & t \geq \frac{(\nu^*)^2}{nb^*} \end{cases} \tag{59}$$

Since $b^* = \max_k b_k$, and we are given that all $b_k = 1$, $b^* = 1$ as well. For ν^*, we have

$$\nu^* = \sqrt{\sum_{k=1}^{n} O(\gamma)^2} = \sqrt{nO(\gamma^2)} \tag{60}$$

$$\Pr\left[\frac{1}{n}X^* \geq t\right] \leq \begin{cases} e^{-\frac{nt^2}{2O(\gamma^2)}} & 0 \leq t < O(\gamma^2) \\ e^{-\frac{nt}{2}} & t \geq O(\gamma^2) \end{cases} \tag{61}$$

Compare the two tail bounds for X^*/n and discuss which one is stronger. Intuitively discuss why one bound is stronger than the other when n is sufficiently big (as γ is fixed.)

First, notice that the only difference between these two tail bounds and those derived in part (c) are the factor of n in the numerator of the fraction in the exponential. Notably, the sub-exponential bound still applies for the same ranges of t as in part (c). Since $X^*/n \xrightarrow{p} 0$, the region of "interest" for t, so to speak, also decreases as n increases. That means that, for large n, the fast regime in $0 \leq t < O(\gamma^2)$ becomes more relevant with large n. Since $\gamma^2 \ll 1$, the sub-exponential bound in the fast regime is stronger than the sub-Gaussian bound.

Problem 2(a)

Let m be the median of P, i.e. m that satisfies $P(X \leq m) = \frac{1}{2}$. Let $\hat{m}_n = med(X_1, \ldots, X_n)$ be the sample median, defined as

$$med(X_1, \ldots, X_n) \triangleq \frac{1}{2}\left(X_{(n/2)} + X_{(n/2+1)}\right) \quad \textit{where } X_{(1)} \leq \cdots \leq X_{(n)} \textit{ are the sorted version of } X_1, \ldots, X_n \tag{62}$$

For any $t > 0$, show that conditioned on the event $\hat{m}_n < m - t$, the following event happens with probability 1

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i < m - t\} \;\geq\; \frac{1}{2} \tag{63}$$

If $\hat{m}_n < m - t$, then we know that $X_i < m - t$ for all $i \geq \frac{n}{2}$ by definition of $\hat{m}_n$. Therefore,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i < m - t\} \geq \frac{1}{n}\sum_{i=1}^{n/2} \mathbf{1}\{X_i < m - t\} = \frac{1}{n}\frac{n}{2} = \frac{1}{2} \tag{64}$$

Problem 2(b)

Let $Y_i = \mathbf{1}\{X_i < m - t\}$, with $\mathbb{E}[Y_i] = P(X_i < m - t)$. Since $Y_i \in \{0, 1\}$, we know that Y_i is sub-Gaussian with parameter $\sigma^2 = \frac{1}{4}$. Therefore, $Z = \sum_i Y_i$ is also sub-Gaussian, and has parameter $\sigma^2 = \sum_{i=1}^n \frac{1}{4} = \frac{n}{4}$. This gives us the following tail bounds:

$$\Pr\left[|Y_i - \mu_{Y_i}| \geq t\right] \leq 2\exp\left(-2t^2\right) \tag{65}$$

$$\Pr\left[|Z - \mu_Z| \geq t\right] \leq 2\exp\left(-2\frac{t^2}{n}\right) \tag{66}$$

Furthermore we can also show that the event $\{\frac{1}{n}Z \geq \frac{1}{2}\}$ can be expressed as an event containing an integral over the density $p(x)$, which we can then use to get a bound in terms of $p(m)$.

$$\begin{aligned}
\Pr\left[\frac{1}{n}Z \geq \frac{1}{2}\right] &= \Pr\left[Z \geq n/2\right] && (67)\\
&= \Pr\left[Z - \mu_z \geq n/2 - \mu_z\right] && (68)\\
\frac{n}{2} - \mu_z &= \frac{n}{2} - n\Pr\left[X < m - t\right] && (69)\\
&= n\left[\frac{1}{2} - \Pr\left[X < m - t\right]\right] && (70)\\
\frac{1}{2} - \Pr\left[X < m - t\right] &= \Pr\left[X \leq m\right] - \Pr\left[X < m - t\right] && (71)\\
&= \int_{m-t}^{m} p(x)\mathrm{d}x && (72)
\end{aligned}$$

We'll now use the integral in 72 to get a bound in terms of $p(m)$. We first note that $\forall \epsilon > 0$, $\exists t_0 > 0$ such that $\forall t' \in (0, t_0)$,

$$|p(m) - p(m - t')| \leq \epsilon \tag{73}$$

Therefore, $\forall \epsilon > 0$, $\exists t_0 > 0$ such that $\forall t' \in (0, t_0)$:

$$\begin{aligned}
\int_{m-t}^{m} p(x)\mathrm{d}x &\geq \int_{m-t}^{m} p(m)\mathrm{d}x - \int_{m-t}^{m} |p(m - t') - p(m)|\mathrm{d}x && (74)\\
&\geq tp(m) - \epsilon t && (75)
\end{aligned}$$

If $\epsilon < \frac{1}{2}p(m)$, this yields $tp(m) - \epsilon t \geq \frac{1}{2}tp(m)$. We can plug this back in to obtain the desired result:

$$\begin{aligned}
\Pr\left[\frac{1}{n}Z \geq \frac{1}{2}\right] &\leq \exp\left(-2n\left(\int_{m-t}^{m} p(x)\mathrm{d}x\right)^2\right) && (76)\\
&\leq \exp\left(-2n\left(\frac{1}{2}tp(m)\right)^2\right) && (77)\\
&= \exp\left(-\frac{1}{2}np(m)^2t^2\right) && (78)
\end{aligned}$$

Problem 2(c)

Show $\exists n_0$ *s.t.* $\forall n \geq n_0$ *w.p. at least 1 -* δ*, we have*

$$|\hat{m}_n - m| \leq \frac{1}{p(m)} \sqrt{\frac{2\log(2/\delta)}{n}} \tag{79}$$

From part(a), we know that the following two events occur with equal probability:

$$\hat{m}_n < m - t \qquad \text{and} \qquad \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i < m - t\} \geq \frac{1}{2} \tag{80}$$

Note that this is true because, if we let $A := \{\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i < m - t\} \geq \frac{1}{2}\}$ and $B := \{\hat{m}_n < m - t\}$, we can see that

$$\begin{aligned}
\Pr[A] &= P(A, B) + P(A, \neg B) && (81)\\
&= P(A \mid B)P(B) + P(A \mid \neg B)P(\neg B) && (82)\\
&= P(B) + P(A \mid \neg B)P(\neg B) \quad \textbf{[part (a)]} && (83)\\
&= P(B) && (84)
\end{aligned}$$

since $\Pr[A \mid \neg B] = 0$. From part (b) we know that

$$\Pr[\hat{m}_n < m - t] \leq \exp\left(-\frac{1}{2}np(m)^2t^2\right) \tag{85}$$

In other words

$$\Pr[|\hat{m}_n - m| \leq t] \geq 1 - 2\exp\left(-\frac{1}{2}np(m)^2t^2\right) \tag{86}$$

Let $t = \frac{1}{p(m)}\sqrt{2\log(2/\delta)/n}$. Then

$$2\exp\left(-\frac{1}{2}np(m)^2t^2\right) = \delta \tag{87}$$

which means the following event occurs with probability at least $1 - \delta$

$$|\hat{m}_n - m| \leq \frac{1}{p(m)} \sqrt{\frac{2\log(2/\delta)}{n}} \tag{88}$$

Problem 2(d)

Show that

$$Y_{(\lfloor n/2-\epsilon n\rfloor)} \leq \hat{m}_n \leq Y_{(\lceil n/1+1\rceil)} \tag{89}$$

We can show this by bounding the number of Y_i's that are larger/smaller than $\hat{m}_n$. Let $N(x) := |\{Y_i : Y_i \geq \hat{x}\}|$ denote the number of Y_i's greater than or equal to x. By definition, we know that there are at least $n/2$ variables in $\{X_1, \ldots, X_n\}$ greater than or equal to the median. Since we are told that $\epsilon > 1/2$, we also know that $(1-\epsilon)n = n - k > n/2$ (we also know that n is even). Therefore, we can assert

$$N(\hat{m}_n) \geq \frac{n}{2} - k \tag{90}$$

If there are at least $\frac{n}{2} - k$ number of Y_i's greater than or equal to $\hat{m}_n$, then necessarily there must be at most $(n-k) - (\frac{n}{2} - k) = \frac{n}{2}$ number of Y_i's strictly less than $\hat{m}_n$. Therefore, $\hat{m}_n \leq Y_{(\lceil n/2+1\rceil)}$.

We can apply the same kind of logic to obtain the other side of the desired inequality. Namely, since at most $n/2$ number of Y_i's are strictly greater than $\hat{m}_n$, we have that at least $(n-k)-\frac{n}{2} = \frac{n}{2} - k = \frac{n}{2} - \epsilon n$ are less than or equal to $\hat{m}_n$. Combining these two results yields

$$Y_{(\lfloor n/2-\epsilon n\rfloor)} \leq \hat{m}_n \leq Y_{(\lceil n/1+1\rceil)} \tag{91}$$

Problem 2(e)

Show $\exists n_0$ *and* $\exists \epsilon_0$ *s.t.* $\forall n \geq n_0$ *and* $\forall \epsilon \leq \epsilon_0$*, w.p. at least* $1-\delta$

$$|\hat{m}_n - \mu| \leq C\left(\epsilon + \sqrt{\log(2/\delta)/n}\right) \tag{92}$$

As suggested by the hint, proceed by proving the concentration of the quantiles. Let

$$\alpha_- = \frac{1-2\epsilon}{2(1-\epsilon)} \tag{93}$$

$$\alpha_+ = \frac{\frac{n}{2}+1}{(1-\epsilon)n} \tag{94}$$

Similarly, let q_{α_-} and q_{α_+} denote the corresponding quantiles of $\mathcal{N}(\mu, 1)$ (the distribution of the Y_i's). We can then apply the result from part (b)

$$\Pr\left[Y_{(\lceil n/2+1\rceil)} - q_{\alpha_+} \geq t\right] \leq \exp\left(-\frac{1}{2}(1-\epsilon)np(q_{\alpha_+})^2t^2\right) \tag{95}$$

$$< \exp\left(-\frac{1}{4}np(q_{\alpha_+})^2t^2\right) \tag{96}$$

where the last inequality follows from the fact that $\epsilon < \frac{1}{2}$. Then we can apply part (c) to assert that there exists n_0 such that for any even integer $n \geq n_0$, with probability at least $1-\delta$, we have

$$Y_{(\lceil n/2+1\rceil)} \leq q_{\alpha_+} + \frac{1}{p(q_{\alpha_+})}\sqrt{\frac{4\log(2/\delta)}{n}} \tag{97}$$

We can repeat the previous steps for $Y_{(n/2-\epsilon n)}$ and α_- as follows:

$$\Pr\left[Y_{(\lfloor n/2-\epsilon n\rfloor)} - q_{\alpha_-} \geq t\right] \leq \exp\left(-\frac{1}{4}np(q_{\alpha_-})^2t^2\right) \tag{98}$$

$$Y_{(\lfloor n/2-\epsilon n\rfloor)} \geq q_{\alpha_-} + \frac{1}{p(q_{\alpha_-})}\sqrt{\frac{4\log(2/\delta)}{n}} \tag{99}$$

So with probability at least $1-\delta$, and using the result from part (d):

$$q_{\alpha_-} + \frac{1}{p(q_{\alpha_-})}\sqrt{\frac{4\log(2/\delta)}{n}} \leq Y_{(\lfloor n/2-\epsilon n\rfloor)} \leq \hat{m}_n \leq Y_{(\lceil n/2+1\rceil)} \leq q_{\alpha_+} + \frac{1}{p(q_{\alpha_+})}\sqrt{\frac{4\log(2/\delta)}{n}} \tag{100}$$

Therefore if we can show, for $n \geq n_0$ and $\epsilon \leq \epsilon_0$, that q_{α_-} and q_{α_+} can be bounded in the form of

$$q_{\alpha_-} \geq \mu - C\epsilon \tag{101}$$

$$q_{\alpha_+} \leq \mu + C\epsilon \tag{102}$$

we'll have the desired result. Note that q_x, which here is the xth quantile of the normal distribution, is continuous and differentiable[2], with derivative being the PDF of the normal distribution. Note that $q_{\frac{1}{2}} = \mu$. We can then do a taylor expansion centered on $q_{\frac{1}{2}}$, with a temporary abuse of notation (the δ below is unrelated to the δ in this problem):

$$q_{\frac{1}{2}-\delta} \geq \mu - C\delta \tag{104}$$

$$q_{\frac{1}{2}+\delta} \leq \mu + C\delta \tag{105}$$

Since $\alpha_+ = \frac{1}{2} + \frac{1}{1-\epsilon}\left(\frac{1}{2}\epsilon + n\right)$,

$$q_{\alpha_+} \leq \mu + C\frac{1}{1-\epsilon}\left(\frac{1}{2}\epsilon + n\right) \tag{106}$$

$$q_{\alpha_+} + \frac{1}{p(q_{\alpha_+})}\sqrt{\frac{4\log(2/\delta)}{n}} \leq \mu + C\epsilon + \frac{1}{p(q_{\alpha_+})}\sqrt{\frac{4\log(2/\delta)}{n}} \tag{107}$$

$$\leq \mu + C\left(\epsilon + \sqrt{\frac{\log(2/\delta)}{n}}\right) \tag{108}$$

for $C > \frac{2}{p(q_{\alpha_+})}$.

Similarly, $\alpha_- = \frac{1}{2} - \frac{1}{1-\epsilon}\left(\frac{1}{2}\epsilon\right)$, which results in

$$q_{\alpha_-} + \frac{1}{p(q_{\alpha_-})}\sqrt{\frac{4\log(2/\delta)}{n}} \geq \mu - C\left(\epsilon + \sqrt{\frac{\log(2/\delta)}{n}}\right) \tag{109}$$

for $C > \frac{2}{p(q_{\alpha_-})}$.

Combining these two results, we require $C > \frac{2}{\min\{p(q_{\alpha_+}), p(q_{\alpha_-})\}}$. Therefore

$$\mu - C\left(\epsilon + \sqrt{\frac{\log(2/\delta)}{n}}\right) \leq \hat{m}_n \leq \mu + C\left(\epsilon + \sqrt{\frac{\log(2/\delta)}{n}}\right) \tag{110}$$

and we have the desired result.

[2]Specifically, for the normal distribution, the quantile function $q_x = F_X^{-1}(x)$ is the inverse CDF. The derivative of q_x is thus the derivative of the inverse CDF of the normal distribution, which is

$$\frac{d}{dx}F_X^{-1}(x) = \frac{1}{p(F_X^{-1}(x))} = \frac{1}{p(q_x)} \tag{103}$$

where $p(x)$ is the density function.