Answer to 3.1.

(a) Let \( p = \mathbb{P}(X \leq x) \). Notice \( nT_n \sim \text{Bin}(n, p) \). By the CLT \( \sqrt{n} (T_n - p) \overset{d}{\to} N(0, p(1 - p)) \).

(b) Notice \( \sqrt{n}(\bar{X}_n - \theta) \sim N(0, 1) \). Let \( G(y) = \Phi(x - y) \). Then \( G'(\theta) = -\phi(x - \theta) \). By the delta method \( \sqrt{n}(G(\bar{X}_n) - G(\theta)) = \sqrt{n}(G_n - p) \overset{d}{\to} N(0, \phi(x - \theta)^2) \).

(c) By the proposition from lecture, the ARE is the inverse ratio of the asymptotic variances:
\[
\text{ARE}(G_n, T_n) = \frac{\frac{p(1-p)}{\phi^2(x-\theta)}}{\text{var}(\hat{\theta}_n)}.
\]
It can be verified that this is always larger than 1.57, so \( G_n \) is a more efficient estimator.

(d) By the LLN we have \( \bar{X} \overset{P}{\to} \mathbb{E}_P[X_1] \). By the continuous mapping theorem and we have \( G_n \overset{P}{\to} \Phi(x - \mathbb{E}_P[X_1]) \). There exists \( x \in \mathbb{R} \) such that \( \Phi(x - \mathbb{E}_P[X_1]) \neq \mathbb{P}(X \leq x) \), or otherwise \( X \) is normally distributed.

(e) We have \( p = \Phi(x - \theta) \), so take \( \hat{\theta}_n = x - \Phi^{-1}(T_n) \). By the LLN \( T_n \overset{P}{\to} p \), so by the continuous mapping theorem \( \hat{\theta}_n \overset{P}{\to} \theta \), so \( \hat{\theta}_n \) is consistent. Now for \( F(t) = x - \Phi^{-1}(t) \), the inverse function theorem says that \( F'(p) = \frac{1}{\phi(x - \theta)} \). Using part (a) and the delta method we have \( \sqrt{n}(\hat{\theta}_n - \theta) = \sqrt{n}(x - \Phi^{-1}(T_n) - (x - \Phi^{-1}(p))) \overset{d}{\to} N(0, \frac{p(1-p)}{\phi^2(x-\theta)^2}) \). Thus \( \text{ARE}(\hat{\theta}_n, \bar{X}_n) = \frac{\phi(x-\theta)^2}{p(1-p)} \). By the remark in the answer to part (c), this is less than 1, and the MLE \( \bar{X}_n \) is more efficient.

Answer to 3.2.

(a) Performing a Taylor expansion on \( \nabla L_n(\theta) \) at \( \theta_0 \), we get
\[
\nabla L_n(\hat{\theta}_n) = \nabla L_n(\theta_0) + R(\hat{\theta}_n - \theta_0),
\]
where \( R = R(X_{1:n}) \in \mathbb{R}^{d \times d} \) satisfy
\[
\left\| R - \nabla^2 L_n(\hat{\theta}_n) \right\|_{op} \leq \left( \frac{1}{n} \sum_{i=1}^{n} M(X_i) \right) \| \hat{\theta}_n - \theta_0 \|_2 = O_{P_{\theta_0}} (\| \hat{\theta}_n - \theta_0 \|_2).
\]
Plugging this into the definition of \( \delta_n \), we get
\[
\nabla L_n(\theta_0) + \left( \nabla^2 L_n(\hat{\theta}_n) + O_{P_{\theta_0}} (\| \hat{\theta}_n - \theta_0 \|_2) \right) (\hat{\theta}_n - \theta_0) + \nabla^2 L_n(\hat{\theta}_n) (\delta_n - \hat{\theta}_n) = 0.
\]
Multiplying both sides by \( \sqrt{n} \) and notice that \( \delta_n - \hat{\theta}_n = \delta_n - \theta_0 + \theta_0 - \hat{\theta}_n \), we get
\[
0 = \sqrt{n} \nabla L_n(\theta_0) + \nabla^2 L_n(\hat{\theta}_n) \cdot \sqrt{n}(\delta_n - \theta_0) + \sqrt{n} \left( \nabla^2 L_n(\hat{\theta}_n) - \nabla^2 L_n(\theta_0) + O_{P_{\theta_0}} (\| \hat{\theta}_n - \theta_0 \|_2) \right) (\hat{\theta}_n - \theta_0)
\]
\[
= \sqrt{n} \nabla L_n(\theta_0) + \nabla^2 L_n(\hat{\theta}_n) \cdot \sqrt{n}(\delta_n - \theta_0) + \frac{1}{\sqrt{n}} O_{P_{\theta_0}} \left( (\sqrt{n}\| \hat{\theta}_n - \theta_0 \|_2)^2 \right).
\]
Rearranging the above equality gives
\[
\sqrt{n}(\delta_n - \theta_0) = \left( \nabla^2 L_n(\hat{\theta}_n) \right)^{-1} \left( \sqrt{n} \nabla L_n(\theta_0) + O_{P_{\theta_0}} \left( \frac{1}{\sqrt{n}} \right) \right).
\]
Finally, applying the known result $\nabla^2 L_n(\hat{\theta}_n) = \nabla^2 L(\theta_0) + o_{P_0}(1)$ (from Lipschitz continuity) and $\sqrt{n}\nabla L_n(\theta_0) \xrightarrow{d} N(0, \text{Cov}(\nabla \ell(\theta_0; X)))$ (from CLT), we get

$$\sqrt{n}(\hat{\delta}_n - \theta_0) \xrightarrow{d} N(0, (\nabla^2 L(\theta_0))^{-1}\text{Cov}(\nabla \ell(\theta_0; X)(\nabla^2 L(\theta_0))^{-1}) = N(0, I(\theta_0)^{-1}).$$

(b) By symmetry, it is no loss of generality to assume that $\theta = 0$, so we wish to show that $\sqrt{n}\hat{\theta}_n = O_P(1)$. Define $F(c) = \Pr(X \leq c)$, so that

$$F(c) = \frac{1}{2} + \frac{1}{\pi} \arctan(c) = \frac{1}{2} + \frac{1}{\pi}(c + O(c^3)),$$

the latter equality holding as $|c| \to 0$. Let $Y_i = 1 \{X_i \leq c\}$, which is Bernoulli with parameter $1 - F(c) = \frac{1}{2} - \frac{c}{\pi} + O(c^3)$. Then $\hat{\theta}_n \geq c$ if and only if $\sum_{i=1}^n Y_i \geq n/2$, and a Hoeffding bound (or any other tail inequality) gives

$$\Pr(\hat{\theta}_n \geq c) \leq \Pr\left(\sum_{i=1}^n Y_i \geq \frac{n}{2}\right) = \Pr\left(\sum_{i=1}^n (Y_i - \mathbb{E}[Y]) \geq \frac{n}{2} - n\mathbb{E}[Y]\right) \leq \exp\left(-n\left(\mathbb{E}[Y] - \frac{1}{2}\right)^2\right) \leq \exp\left(-n\frac{c^2 + O(c^4)}{\pi^2}\right).$$

In particular, if we set $c = M/\sqrt{n}$ for some $M < \infty$, we have that

$$\lim_{n \to \infty} \Pr(\hat{\theta}_n \geq M/\sqrt{n}) \leq \exp\left(-\frac{M^2}{\pi}\right).$$

A similar calculation gives the lower bound on $\hat{\theta}_n$ and we have $\sqrt{n}\hat{\theta}_n = O_P(1)$ as desired.

(c) We need only compute the Fisher information $I(\theta_0)$. WLOG let $\theta_0 = 0$. We have

$$\ell(\theta; x) = -\log(1 + (x - \theta)^2) + C.$$

Taking the derivative gives

$$\ell'(\theta; x) = \frac{2(x - \theta)}{1 + (x - \theta)^2}.$$ 

So the Fisher information equals

$$I(\theta) = \text{Cov}(\ell'(0; X)) = \text{Cov}\left(\frac{2X}{1 + X^2}\right) = \mathbb{E}\left[\frac{4X^2}{(1 + X^2)^2}\right] = \int_{-\infty}^{\infty} \frac{4x^2}{\pi(1 + x^2)^3} dx.$$

This integral can be computed via a trigonometric transform: let $x = \tan u$, $u \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and so $u = \arctan x$ and $du = \frac{1}{1 + x^2} dx$. Notice that $\frac{1}{1 + \tan^2 u} = \cos^2 u$. We get

$$\int_{-\infty}^{\infty} \frac{4x^2}{\pi(1 + x^2)^3} dx = \int_{-\pi/2}^{\pi/2} \frac{4 \tan^2 u}{\pi(1 + \tan^2 u)^2} du = \frac{4}{\pi} \int_{-\pi/2}^{\pi/2} \sin^2 u \cos^2 u du = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \sin^2(2u) du = \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} (1 - \cos(4u)) du = \frac{1}{2\pi} \cdot \pi = \frac{1}{2}.$$

So the asymptotic distribution of $\hat{\delta}_n$ is $\sqrt{n}(\hat{\delta}_n - \theta_0) \xrightarrow{d} N(0, 2)$. 

2
Answer to 3.3: Here we see that the risk of the Hodges estimator much large than the MLE for values of \( \mu \) that are near zero, although there is an improvement near zero.

Note: the above figures are courtesy of Kenneth Tay.

Answer to 5.1: For distributions \( P \) and \( Q \), we have

\[
d^2_{\text{hel}}(P^n, Q^n) = 1 - (1 - d^2_{\text{hel}}(P, Q))^n.
\]

Now, let \( P = N(\mu, \sigma^2) \) and \( Q = N(0, \sigma^2) \), so that

\[
d^2_{\text{hel}}(P, Q) = 1 - \int_{-\infty}^{\infty} \sqrt{p(x)q(x)} \, dx
\]

\[
= 1 - \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp \left( -\frac{1}{4\sigma^2} x^2 - \frac{1}{4\sigma^2} (x - \mu)^2 \right) \, dx
\]

\[
= 1 - \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp \left( -\frac{1}{2\sigma^2} (x - \mu/2)^2 - \frac{\mu^2}{8\sigma^2} \right) \, dx
\]

\[
= 1 - e^{-\frac{\mu^2}{8\sigma^2}}.
\]

Thus, we have for \( \theta_0 = (0, \sigma^2) \) and \( \theta = (\mu, \sigma^2) \) that

\[
P_{\theta}(T_n = 1) \leq P_{\theta_0}(T_n = 1) + \| P^n_{\theta} - P^n_{\theta_0} \|_{\text{TV}} \leq \alpha + \sqrt{2d_{\text{hel}}(P^n_{\theta}, P^n_{\theta_0})} = \alpha + \sqrt{2} \sqrt{1 - \exp \left( -\frac{\mu^2}{8\sigma^2} \right)}.
\]

So for any given \( \mu \in \mathbb{R}, \mu \neq 0 \), and \( \epsilon > 0 \), we solve

\[
1 - \exp \left( -\frac{n\mu^2}{8\sigma^2} \right) = \frac{\epsilon^2}{2} \quad \text{or} \quad \sigma^2 \geq \frac{n\mu^2}{8|\log(1 - \epsilon^2/2)|}.
\]
That is, roughly, any $\sigma^2 \geq \frac{n\mu^2}{4\epsilon^2}$ is sufficient to guarantee that $\pi_n(\theta) \leq \alpha + \epsilon$.

For the second result, any standard test will work. For example, let $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ be the sample variance. Then the test

$$
\psi_n(X) := \begin{cases} 
1 & \text{if } |X_n| \geq \hat{\sigma} z_{1-\alpha/2} \\
0 & \text{otherwise}
\end{cases}
$$

for $z_{1-\alpha/2}$ defined by $\mathbb{P}(|Z| \geq z_{1-\alpha/2}) = \alpha$ for $Z \sim \mathcal{N}(0,1)$ works. \qed