Let $X_1, \cdots, X_n \overset{iid}{\sim} N(\Theta, \sigma^2)$, with σ^2 known. Let $\Theta \sim N(\mu, b^2)$

$$\begin{aligned}
\pi(\theta \mid X) &\propto \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(X_i-\theta)^2\right\} \times \frac{1}{\sqrt{2\pi b^2}} \exp\left\{-\frac{1}{2b^2}(\theta-\mu)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\theta)^2 - \frac{1}{2b^2}(\theta-\mu)^2\right\} \\
&\propto \cdots \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{b^2}\right)\theta^2 + \left(\frac{n\bar{X}}{\sigma^2}+\frac{\mu}{b^2}\right)\theta\right\} \\
&\propto \exp\left\{-\frac{1}{2\tilde{\sigma}^2}(\theta-\tilde{\mu})^2\right\}.
\end{aligned}$$

The posterior distribution of Θ given X is $N(\tilde{\mu}, \tilde{\sigma}^2)$ where

$$\tilde{\mu} = \frac{n\bar{X}/\sigma^2 + \mu/b^2}{n/\sigma^2 + 1/b^2} \quad \text{and} \quad \tilde{\sigma}^2 = \frac{1}{n/\sigma^2 + 1/b^2}$$

Hence, the posterior mean of $\Theta \mid X$ is $\frac{n\bar{X}/\sigma^2 + \mu/b^2}{n/\sigma^2 + 1/b^2}$ and similarly we can rewrite as

$$\underbrace{\frac{n/\sigma^2}{n/\sigma^2 + 1/b^2}}_{1 \text{ as } n\to\infty} \bar{X} + \underbrace{\frac{1/b^2}{n/\sigma^2 + 1/b^2}}_{0 \text{ as } n\to\infty} \mu$$

Thus, Bayes estimator δ_Λ is $\tilde{\mu}$ if we adopt the squared loss function.

Example 4 (Bayes estimator of weighted L^2 loss) *Assume that we consider $L(\theta, d) = \omega(\theta)\{d - g(\theta)\}^2$, where $\omega(\theta) \geqslant 0$, which can be interpreted as a weight function. Our goal is to find the corresponding Bayes estimator, which minimizes $E(\omega(\Theta)\{g(\Theta) - d\}^2 \mid X = x)$ (*)with respect to d.*
(*) can be rewritten as

$$d^2E(\omega(\Theta) \mid X = x) - 2dE(\omega(\Theta)g(\Theta) \mid X = x) + E(\omega(\Theta)g(\Theta)^2 \mid X = x). \qquad (\dagger)$$

Tanking derivative of (†) with respect to d, we obtain

$$2d^* E(\omega(\Theta) \mid X = x) - 2E(\omega(\Theta)g(\Theta) \mid X = x) = 0.$$

Thus

$$\delta_\Lambda(x) = d^* = \frac{E(\omega(\Theta)g(\Theta) \mid X = x)}{E(\omega(\Theta) \mid X = x)}. \qquad (5)$$

In particular, if $\omega(\cdot) \equiv 1$, $\delta_\Lambda(x)$ (with $\omega(\cdot) \equiv 1$)$=E(g(\Theta) \mid X = x)$.

Theorem 2 *If δ is unbiased for $g(\theta)$ with $r(\Lambda, \delta) < \infty$ and $E(g(\Theta)^2) < \infty$, then δ is not Bayes under the squared loss function unless its average risk is zero, which is*

$$E_{(X,\Theta)}(\{\delta(X) - g(\Theta)\}^2) = 0. \qquad (6)$$

Proof 2 *Let δ be an unbiased estimator under the squared loss function. Then we know that δ is the posterior mean, which is*

$$\delta(X) = E(g(\Theta) \mid X),$$

almost surely. Thus, we have

$$\begin{aligned}
E(\delta(X)g(\Theta)) &= E(E(\delta(X)g(\Theta) \mid X)) \\
&= E(\delta(X)E(g(\Theta) \mid X)) \\
&= E(\delta^2(X)). \qquad (7
\end{aligned}$$

Also,

$$\begin{aligned}
E(\delta(X)g(\Theta)) &= E(E(\delta(X)g(\Theta) \mid \Theta)) \\
&= E(g(\Theta)E(\delta(X) \mid \Theta)) \\
&= E(g^2(\Theta)). \qquad (8
\end{aligned}$$

Observe that

$$\begin{aligned}
E(\{\delta(X) - g(\Theta)\}^2) &= E(\delta^2(X)) - 2E(\delta(X)g(\Theta)) + E(g^2(\Theta)) \\
&= E(\delta^2(X)) - E(\delta(X)g(\Theta)) + E(g^2(\Theta)) - E(\delta(X)g(\Theta)) \\
&= E(\delta^2(X)) - E(\delta^2(X)) + E(g^2(\Theta)) - E(g^2(\Theta)) \text{ (due to (7) and (8))} \\
&= 0.
\end{aligned}$$

Theorem 2 (TPE 5.1.4)**.** *Suppose δ_Λ is Bayes for Λ with*

$$r_\Lambda = \sup_\theta R(\theta, \delta_\Lambda)$$

i.e. the Bayes risk of δ_Λ is the maximum risk of δ_Λ, then:

(i) δ_Λ is minimax,

(ii) Λ is a least favorable prior,

(iii) If δ_Λ is the unique Bayes estimator for Λ almost surely, for all P_θ, then it is a unique minimax estimator.

Proof. (i) Let δ be any other estimator, then we have that:

$$\sup_{\theta\in\Omega} R(\theta, \delta) \geq \int R(\theta, \delta) d\Lambda(\theta) \overset{(*)}{\geq} \int R(\theta, \delta_\Lambda) d\Lambda(\theta)$$

This implies that δ_Λ is minimax.

(ii) If $\delta\Lambda$ is the unique Bayes estimator, then the ineqality above (*) is strict for $\delta \neq \delta_\Lambda$, which implies that δ_Λ is the unique minimax.

(iii) Let Λ' be any other prior distribution, then

$$\begin{aligned}
r_{\Lambda'} &\leq \inf_\delta \int R(\theta, \delta) d\Lambda'(\theta) \leq \int R(\theta, \delta_\Lambda) d\Lambda'(\theta) \\
&\leq \sup_\theta R(\theta, \delta_\Lambda) = r_\Lambda
\end{aligned}$$

1. **Reduce the composite alternative to a simple alternative**: If H_1 is composite, fix $\theta_1 \in \Omega_1$, and test the null hypothesis against the simple alternative $\theta = \theta_1$. (Hope that doesn't depend on θ_1.)

2. **Collapse the composite null to a simple null**: If H_0 is composite, collapse the null hypothesis to a simple one by averaging over the null space Ω_0. We will discuss this strategy in today's lecture.

3. **Apply Neyman Pearson lemma**: Find the MP LRT for testing the resulting simple null versus the resulting simple alternative using the NP lemma. Note that if the resulting test does not depend on θ_1, then it will be UMP for the H_0 vs H_1.

Example 1. *Suppose $X \sim Binomial(n, \theta)$ for some $\theta \in (0, 1)$ and we adopt the squared loss function, is $\frac{x}{n}$ minimax?*

Notice that the corresponding risk is $R(\theta, \frac{x}{n}) = \frac{\theta(1-\theta)}{n}$. Observe that the risk has a unique maximum at $\theta = \frac{1}{2}$. The worst risk is:

$$\sup_{\theta\in\Omega} R(\theta, \frac{x}{n}) = R(\frac{1}{2}, \frac{1}{2}) = \frac{1}{4n}$$

In this case, [TPE 5.1.6] is not helpful because if $\Lambda(\{\frac{1}{2}\}) = 1$, then $\delta_\Lambda(X) = \frac{1}{2} \neq \frac{x}{n}$.

However, [TPE 5.1.5] can be helpful instead. To find a minimax estimator, we will need to search for a prior such that the Bayes estimator has constant risk.

Recall that if the prior is $Beta(\alpha, \beta)$, the Bayes estimator under the squared loss is:

$$\delta_{\alpha,\beta}(X) = \frac{x + \alpha}{n + \alpha + \beta}$$

for any α, β.

$$\begin{aligned}
R(\theta, \delta_{\alpha,\beta}) &= \mathbb{E}_\theta\left(\left\{\frac{x+\alpha}{n+\alpha+\beta} - \theta\right\}^2\right) \\
&= \frac{1}{(n+\alpha+\beta)^2}\mathbb{E}_\theta\left(\{x - n\theta - \alpha(\theta-1) - \theta\beta\}^2\right) \\
&= \frac{1}{(n+\alpha+\beta)^2}\left[n\theta(1-\theta + \{\alpha(\theta-1) + \theta\beta\}^2)\right]
\end{aligned}$$

To eliminate the θ dependence in $R(\theta, \delta_{\alpha,\beta})$, we need to set the coefficients of θ^2 and θ be zero, that is:

$$\begin{aligned}
-n + (\alpha+\beta)^2 &= 0 \\
n - 2\alpha(\alpha+\beta) &= 0,
\end{aligned}$$

which solves $\alpha = \beta = \frac{\sqrt{n}}{2}$. The Bayes estimator $\delta_{\frac{\sqrt{n}}{2},\frac{\sqrt{n}}{2}}(X) = \frac{X+\sqrt{n}/2}{n+\sqrt{n}}$ is minimax (TPE 5.1.4) with constant risk of $\frac{1}{4(\sqrt{n}+1)^2}$, we can conclude that $\frac{X}{n}$ is not minimax.

Theorem 7 (TPE 5.1.12)**.** *Suppose there is a real number r such that $\{\Lambda_m\}$ is a sequence of priors with $r_{\Lambda_m} \to r < \infty$. Let δ be any estimator such that $\sup_\theta R(\theta, \delta) = r$. Then we have*

(i) δ is minimax;

(ii) $\{\Lambda_m\}$ is least-favourable.

Proof. (i) Let δ' be any other estimator. Then for any m, we have

$$\sup_\theta R(\theta, \delta') \geq \int_\Omega R(\theta, \delta') \, d\Lambda_m(\theta) \geq r_{\Lambda_m}.$$

Then sending $m \to \infty$ yields

$$\sup_\theta R(\theta, \delta') \geq r = \sup_\theta R(\theta, \delta),$$

which implies that δ is minimax.

(ii) Let Λ' be any prior, then

$$r_{\Lambda'} = \int_\Omega R(\theta, \delta_{\Lambda'}) \, d\Lambda'(\theta) \leq \int_\Omega R(\theta, \delta) \, d\Lambda'(\theta) \leq \sup_\theta R(\theta, \delta) = r,$$

which means that $\{\Lambda_m\}$ is least favourable.

Example 3 (cont'd)**.** *If we manage to find a sequence of priors $\{\Lambda_m\}$ such that $r_{\Lambda_m} \to \frac{\sigma^2}{n} = r$, then we can obtain a minimax estimator for θ. Let consider the sequence of priors $\Lambda_m \sim \mathcal{N}(0, m^2)$ (Λ_m will tend to the uniform prior over $\mathbb{R}$ which is improper with $\pi(\theta) = 1$ for any $\theta \in \mathbb{R}$). This will yield the following posterior distribution.*

$$\begin{aligned}
f(\theta|x_1, ..., x_n) &\propto \pi(\theta) \cdot f(x_1, ..., x_n|\theta) \\
&\propto \exp\left(-\frac{\theta^2}{2m^2} - \frac{\sum_{i=1}^n (x_i - \theta)^2}{2\sigma^2}\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\frac{1}{m^2} + \frac{n}{\sigma^2}\right)\theta^2 + \frac{n\bar{x}}{\sigma^2} \cdot \theta\right) \\
&\sim \mathcal{N}\left(\frac{\frac{n\bar{x}}{\sigma^2}}{\frac{1}{m^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{m^2} + \frac{n}{\sigma^2}}\right)
\end{aligned}$$

Note that the posterior variance does not depend on $(X_1, ..., X_m)$, hence

$$r_{\Lambda_m} = \frac{1}{\frac{1}{m^2} + \frac{n}{\sigma^2}} \to \frac{\sigma^2}{n} = \sup_\theta R(\theta, \overline{X}).$$

It now follows from Theorem 5.1.12 (TPE) that $\overline{X}$ is minimal and $\{\Lambda_m\}$ is least favourable.

Example 4 (weighted squared loss)**.** *Let $X \sim$ Binomial(n, θ) with the loss function $L(\theta, d) = \frac{(d-\theta)^2}{\theta(1-\theta)}$. We may view this loss function as the weighted squared loss function with weights $w(\theta) = \frac{1}{\theta(1-\theta)}$.*

Note that for any θ, $R(\theta, X/n) = \frac{1}{n}$, which is constant in θ. This suggests that X/n can be minimax.
But be reminded that we cannot directly apply TPE 4.2.3 because L is not the vanilla squared loss function.

Consider the prior $\Theta \sim \Lambda_{\alpha,\beta} =$ Beta(α, β), for some $\alpha, \beta > 0$. By results in Lecture 8, we have $\Theta|X \sim$ Beta$(X + \alpha, n - X + \beta)$ and we can find the Bayes estimator as

$$\delta_\Lambda(X) = \frac{\mathbb{E}_{\Theta|X}\left(\frac{1}{1-\Theta}\middle|X\right)}{\mathbb{E}_{\Theta|X}\left(\frac{1}{\Theta(1-\Theta)}\middle|X\right)}$$

Suppose we have observed $X = x$ with $\alpha + x > 1$ and $n + \beta + x > 1$, then the resulting Bayes estimator is

$$\delta_{\alpha,\beta}(x) = \frac{\alpha + x - 1}{\alpha + \beta + n - 2}.$$

In particular, when $\alpha = \beta = 1$, we have $\delta_{1,1}(x) = x/n$ minimizes posterior risk under prior $\Lambda_{1,1}$ after observing $0 < x < n$.

When $x \in \{0, n\}$, then the posterior risk under the prior $\Lambda_{1,1}$ after observing $X = x$ and deciding $\delta(x) = d$ is

$$\int_0^1 \frac{(d-\theta)^2}{\theta(1-\theta)} \cdot \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)} \cdot \theta^x(1-\theta)^{n-x} \, d\theta,$$

which for $x = 0$ reduces to $\int_0^1 \frac{(n+1)(1-\theta)^{n-1}(d-\theta)^2}{\theta} d\theta$. Note this converges only when $\delta(0) = 0$. Similarly, one can deduce that $\delta(n) = 1$.

Now we may conlcude that X/n minimizes the posterior risk under prior distribution $\Lambda_{1,1}$ for any outcome X. Hence X/n is indeed minimax under such weighted squared loss function.

Definition 3 *We say that the family of densities $\{p_\theta : \theta \in R\}$ has monitone likelihood ratio in $T(x)$ if*

1. $\theta \neq \theta'$ implies $p_\theta \neq p_{\theta'}$ (identifiability)

2. $\theta < \theta'$ implies $p_{\theta'}(x) \,/\, p_\theta(x)$ is a non-decreasing function of $T(x)$ (Monotonicity)

Example 1 *Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\theta, \sigma^2)$, where σ^2 is known, and the parameter θ is the estimand. The minimax estimator is $\bar{X}$ under the squared error loss function.*

Question: is $\bar{X}$ admissible?
[A more general question: when is $a\bar{X} + b, a, b \in \mathbb{R}$ (any affine function of $\bar{X}$) admissible?]

Case 1 *$0 < a < 1$:*
Observe that $a\bar{X} + b$ is a convex combination of $\bar{X}$ and b. It is a Bayes estimator with respect to some Gaussian prior on θ. Since we are considering the squared error loss function, which is strictly convex, the Bayes estimator is unique. By Theorem 5.2.4 (TPE), $a\bar{X} + b$ is therefore admissible.

Case 2 *$a = 0$:*
In this case, b is also a unique Bayes estimator with respect to a degenerate prior distribution with unit mass at $\theta = b$. ($\Lambda(\theta) = N(b, 0^2)$). So by Theorem 5.2.4, b is admissible.

Case 3 *$a = 1, b \neq 0$:*
In this case, $\bar{X} + b$ is not admissible, because it is dominated by $\bar{X}$ because $\bar{X}$ has the same variance as $\bar{X} + b$, but is has a strictly smaller bias.

Case 4 *$a > 1$:*

$$\text{Risk of } a\bar{X} + b = \mathbb{E}[(a\bar{X} + b - \theta)^2] = \mathbb{E}[(a(\bar{X} - \theta) + b + \theta(a-1))^2] = \frac{a^2\sigma^2}{n} + (b + \theta(a-1))^2$$

So, when $a > 1$,

$$\mathbb{E}[(a\bar{X} + b - \theta)^2] \geq \frac{a^2\sigma^2}{n} > \frac{\sigma^2}{n} = R(\theta, \bar{X})$$

Hence, $\bar{X}$ dominates $a\bar{X} + b$ when $a > 1$, and so in this case $a\bar{X} + b$ is inadmissible.

Case 5 *$a < 0$:*

$$\mathbb{E}[(a\bar{X} + b - \theta)^2] > (b + \theta(a-1))^2 = (a-1)^2(\theta + \frac{b}{a-1})^2 > (\theta + \frac{b}{a-1})^2$$

and this is the risk of predicting the constant $-\frac{b}{a-1}$. So $-\frac{b}{a-1}$ dominates $a\bar{X} + b$. Hence, $a\bar{X} + b$ is inadmissible.

Case 6 *$a = 1, b = 0$:*
We use a limiting Bayes argument. Suppose $\bar{X}$ is inadmissible. WLOG, we assume that $\sigma^2 = 1$, and have

$$R(\theta, \bar{X}) = \frac{1}{n}$$

By our hypothesis, there must exist an estimator δ' such that $R(\theta, \delta') \leq \frac{1}{n}$ for all θ, and $R(\theta', \delta') < \frac{1}{n}$ for some $\theta' \in \Omega$ [at least one]. Because $R(\theta, \delta)$ is continuous in θ, there must exist $\epsilon > 0$ and an interval (θ_0, θ_1) containing θ' such that

$$R(\theta, \delta') < \frac{1}{n} - \epsilon, \forall\theta \in (\theta_0, \theta_1) \qquad (*)$$

Let r'_τ be the average risk of δ' with respect to the prior distribution $N(0, \tau^2)$ on θ. Let r_τ be the average risk of a Bayes estimator δ_τ under the same prior.
Note that $\delta_\tau \neq \delta'$ because $R(\theta, \delta_\tau) \to \infty$ as $\theta \to \infty$, which is not consistent with $R(\theta, \delta') \leq \frac{1}{n}$ for all $\theta \in \Omega = \mathbb{R}$. So, $r_\tau < r'_\tau$ because the Bayes estimator is unique almost surely w.r.t. the marginal distribution of θ.

$$\frac{\frac{1}{n} - r'_\tau}{\frac{1}{n} - r_\tau} = \frac{\frac{1}{\sqrt{2\pi\tau^2}}\int_{-\infty}^{\infty}\{\frac{1}{n} - R(\theta, \delta')\}\exp(-\frac{\theta^2}{2\tau^2})d\theta}{\frac{1}{n} - \frac{1}{n + \frac{1}{\tau^2}}} \qquad (\#)$$

By (*), we can simplify (#) as follows

$$\frac{\frac{1}{n} - r'_\tau}{\frac{1}{n} - r_\tau} \geq \frac{\frac{1}{\sqrt{2\pi\tau^2}}\int_{\theta_0}^{\theta_1} \epsilon e^{-\frac{\theta^2}{2\tau^2}} d\theta}{\frac{1}{n(1+n\tau^2)}} = \frac{n(1+n\tau^2)}{\tau\sqrt{2\pi}}\epsilon \int_{\theta_0}^{\theta_1} e^{-\frac{\theta^2}{2\tau^2}} d\theta$$

As $\tau \to \infty$, the first expression $\frac{n(1+n\tau^2)\epsilon}{\tau\sqrt{2\pi}} \to \infty$, and since the integrand converges monotonically to 1, Lebesgue's monotone convergence theorem ensures that the integrand approaches to the quantity $\theta_1 - \theta_0$. So, for sufficiently large τ, we must have

$$\frac{\frac{1}{n} - r'_\tau}{\frac{1}{n} - r_\tau} > 1$$

This means that $r'_\tau < r_\tau$. But this is a contradiction because r_τ is the optimal average risk. So our assumption that there was a dominating estimator is incorrect, in which case $a\bar{X} + b = \bar{X}$ is admissible.

Lemma 1 (Neyman-Pearson Lemma) .

(i) Existence. For testing $H_0 : p_0$ vs $H_1 : p_1$, there exists a test $\phi(X)$ and a constant k such that

(a) $E_{p_0}(\phi(X)) = \alpha$ (size = level)

(b)

$$\phi(x) = \begin{cases} 1, & \text{if } \frac{p_1(x)}{p_0(x)} > k \text{ [Rejection]} \\ 0, & \text{otherwise [Acceptance]} \end{cases}$$

such a test is called a likelihood ratio test.

(ii) Sufficiency: If a test satisfies (a) and (b) for some constant k, it is most powerful for testing $H_0 : p_0$ vs $H_1 : p_1$ at level α.

(iii) Necessity: If a test ϕ is MP at level α, then it satisfies (b) for some k, and it also satisfies (a) unless there exists a test if size $< \alpha$ with power 1.

Example 2 *Let $X_1, ..., X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$ with σ^2 known. We want to test:*

$$H_0 : \mu = 0 \quad vs \quad H_1 : \mu = \mu_1$$

where μ_1 is given. We calculate the likelihood ratio:

$$r(x) = \frac{p_1(x)}{p_0(x)} = \cdots = \exp\left(\frac{1}{\sigma^2}\mu_1 \sum_{i=1}^n x_i - \frac{n\mu_1^2}{2\sigma^2}\right)$$

Suppose

$$\begin{aligned}
r(x) > k &\iff \mu_1 \frac{\sum_{i=1} x_i}{\sigma^2} - \frac{n\mu_1^2}{2\sigma^2} > \log k \\
&\iff \mu_1 \sum_{i=1}^n x_i > k' \\
&\iff \begin{cases} \sum_{i=1}^n x_i > k'', & \text{if } \mu_1 > 0 \\ \sum_{i=1}^n x_i < k''', & \text{if } \mu_1 < 0 \end{cases}
\end{aligned}$$

Let's focus on the case where $\mu_1 > 0$,

$$\begin{aligned}
r(x) > k &\iff \sum_{i=1}^n x_i > k'' \\
&\iff \frac{\sqrt{n}\bar{x}}{\sigma} > k''''
\end{aligned}$$

To calculate k''''(critical value), we need to evaluate

$$\mathbb{E}_{p_0}(\phi(X)) = \alpha = P_{\mu=0}\left(\frac{\bar{X}}{\sigma/\sqrt{n}} > k''''\right)$$

where $\frac{\bar{X}}{\sigma/\sqrt{n}}$ is normally distributed with zero mean.

Proof 1 *Let $\gamma(x) = \frac{p_1(x)}{p_0(x)}$ be the likelihood ratio. Denote the cumulative distribution function of $r(x)$ under H_0 as F_0.*

1. Existence. Let $\alpha(c) = P_0(\gamma(X) > c) = 1 - F_0(c)$ for $c \in R$. Then $\alpha(c)$ is a non-increasing, right-continuous function of c [$\alpha(c^-) = \lim_{\epsilon\to 0}\alpha(c + \epsilon)$]. Observe also that $\alpha(c)$ is not always left-continuous at every value of c, but the left-hand limits exist. There exists a value c_0 such that $\alpha(c_0) \leq \alpha \leq \alpha(c_0^-)$. Note that $F(c_0) \geq 1 - \alpha \geq F(c_0^-)$, ie. c_0 is the $1 - \alpha$ quantile of $r(x)$. we define our test function to be:

$$\phi(x) = \begin{cases} 1 & \text{if } \gamma(x) > c_0, \\ \gamma & \text{if } \gamma(x) = c_0, \\ 0 & \text{if } \gamma(x) < c_0, \end{cases}$$

for some constant γ. The test function always rejects the null if the $\gamma(x) > c_0$, and does not reject the null if $\gamma(x) < c_0$.
The size of the test ϕ is given by

$$\begin{aligned}
E_0(\phi(X)) &= P_0(\gamma(X) > c_0) + \gamma P_0(\gamma(X) = c_0) \\
&= \alpha(c_0) + \gamma(\alpha(c_0^-) - \alpha(c_0))
\end{aligned}$$

If $\alpha(c_0) = \alpha(c_0^-)$ [continuous case], then $\alpha(c_0) = \alpha$ and we automatically have $E_0(\phi(X)) = \alpha$ for any choice of γ. Otherwise, we can set

$$\gamma = \frac{\alpha - \alpha(c_0)}{\alpha(c_0^-) - \alpha(c_0)}$$

which gives also $E_0(\phi(X)) = \alpha$.

2. Sufficiency: Let ϕ satisfies (a) and (b) in part(1), and let ϕ' be any other level α test, which satisfies

$$E_0(\phi'(X)) = \int \phi'(x)p_0(x)d\mu(x) \leq \alpha$$

We need to bound the power difference $E_1(\phi(X)) - E_1(\phi'(X))$ from below by the size difference $E_0(\phi(X)) - E_0(\phi'(X))$ up to a constant multiple.
we claim the following inequality holds:

$$\int (\phi(x) - \phi'(x))(p_1(x) - kp_0(x))d\mu(x) \geq 0$$

To see this, we consider the following three cases:

(a) if $p_1(x) > kp_0(x)$ [$\equiv \gamma(x) > k$], then $\phi(x) = 1$. Since $\phi'(x) \leq 1$, the the integrand is non-negative.

(b) $p_1(x) < kp_0(x)$, ...

(c) $p_1(x) = kp_0(x)$, ...

It implies that

$$\int (\phi(x) - \phi'(x))p_1(x)d\mu(x) \geq k \int (\phi(x) - \phi'(x))p_0(x)d\mu(x) \geq 0$$

meaning that $E_1(\phi(x)) > E_1(\phi'(X))$, ie. ϕ is most powerful at level α.

3. Necessity: Suppose ϕ^* is most powerful at level α, and let ϕ be a likelihood ratio test satisfying (a) and (b), we want to show that $\phi^*(x) = \phi(x)$ except possibly cases where $\frac{p_1(x)}{p_0(x)} = k$ for μ-a.l. x. Define

$$\begin{aligned}
S^+ &= \{x : \phi(x) > \phi^*(x)\} \\
S^- &= \{x : \phi(x) < \phi^*(x)\} \\
S_0 &= \{x : \phi(x) = \phi^*(x)\}
\end{aligned}$$

and $S = (S^+ \cup S^-) \cap \{x : p_1(x) \neq kp_0(x)\}$. We want to show that $\mu(x) = 0$. Suppose $\mu(x) > 0$, as in part 2, we have $(\phi(x) - \phi'(x))(p_1(x) - kp_0(x)) > 0$ on S. Therefore,

$$\begin{aligned}
\int_x (\phi(x) - \phi^*(x))(p_1(x) - kp_0(x))d\mu x &= \int_{S^+\cup S^-} (\phi(x) - \phi^*(x))(p_1(x) - kp_0(x))d\mu x \\
&= \int_S (\phi(x) - \phi^*(x))(p_1(x) - kp_0(x))d\mu x > 0
\end{aligned}$$

By hypothesis, $E_0(\phi(x)) = \alpha$ and $E_0(\phi^*(x)) \leq \alpha$, so the previous inequality implies that

$$E_1(\phi(x)) - E_1(\phi^*(x)) > k(E_0(\phi(x)) - E_0(\phi^*(x)))$$

$\Rightarrow E_1(\phi(x)) \geq E_1(\phi^*(x))$, which contradicts the assumption that ϕ^* is most powerful. Hence $\mu(S) = 0$. It remains to show that the size of ϕ^* is α unless there exists a test of size which is strictly less than α and power 1. For this, note that if size $< \alpha$ and power < 1, we can add points to rejection region until either the size $= \alpha$ or the power is 1.

Theorem 4 (TSH 3.4.1) *Suppose $X \sim p_\theta(x)$ has MLR is $T(x)$ and we test $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$. Then,*

1. there exists a UMP test at level α of the form

$$\phi(x) = \begin{cases} 1 & \text{if } T(x) > k \\ \gamma & \text{if } T(x) = k \\ 0 & \text{if } T(x) < k, \end{cases}$$

where k and γ are determined by $E_{\theta_0}\phi(x) = \alpha_j$.

2. the power function $\beta(\theta) = E_\theta\phi(X)$ is strictly increasing when $0 < \beta(\theta) < 1$.

$$\begin{aligned}
H_0 &: X \sim f_\theta, \ \theta \in \Omega_0 \\
H_1 &: X \sim g,
\end{aligned}$$

where g is known. We now impose a prior distribution Λ on Ω_0. So we consider the new hypothesis

$$H_\Lambda : X \sim h_\Lambda(x) = \int_{\Omega_0} f_\theta(x) \, d\Lambda(\theta),$$

Theorem 1 (TSH 3.8.1)**.** Suppose ϕ_Λ is a MP level-α test for testing H_Λ against g. If ϕ_Λ is level-α for the original hypothesis H_0 (i.e., $\mathbb{E}_{\theta_0}\phi_\Lambda(x) \leq \alpha$, $\forall \theta \in \Omega_0$), then

1. The test ϕ_Λ is MP for original $H_0 : \theta \in \Omega_0$ vs. g.

2. The distribution Λ is least favorable.

Proof. 1. Let ϕ^* be any other level-α test of $H_0 : \theta \in \Omega_0$ vs. g. Then ϕ^* is also a level-α test for H_Λ vs. g, because

$$\mathbb{E}_\theta\phi^*(X) = \int \phi^*(x)f_\theta(x) \, d\mu(x) \leq \alpha, \ \forall \theta \in \Omega_0,$$

which implies that

$$\int \phi^*(x)h_\Lambda(x) \, d\mu(x) = \int\int \phi^*(x)f_\theta(x) \, d\mu(x)d\Lambda(\theta) \leq \int \alpha d\Lambda(\theta) = \alpha.$$

Since ϕ_Λ is MP for H_Λ vs. g, we have

$$\int \phi^*(x)g(x) \, d\mu(x) \leq \int \phi_\Lambda(x)g(x) \, d\mu(x),$$

Hence ϕ_Λ is a MP test for H_0 vs. g, because ϕ_Λ is also level α.

2. Let Λ' be any distribution on Ω_0. Since $\mathbb{E}_\theta\phi_\Lambda(X) \leq \alpha$, $\forall \theta \in \Omega_0$, we know that ϕ_Λ must be level-α for $H_{\Lambda'}$ vs. g. Thus $\beta_\Lambda \leq \beta_{\Lambda'}$, so Λ is least favorable distribution.

Example 1 (Testing in the presence of nuisance parameters)**.** Let $X_1, \ldots, X_n$ be i.i.d. $\mathcal{N}(\theta, \sigma^2)$, where both θ, σ^2 are unknown. We consider testing $H_0 : \sigma \leq \sigma_0$ against $H_1 : \sigma > \sigma_0$. To find a UMP test, we follow the previously mentioned strategy:

1. First we fix a simple alternative (θ_1, σ_1) for some arbitrary θ_1 and $\sigma_1 > \sigma_0$.

2. Second, we choose a prior distribution Λ to collapse our null hypothesis over θ. Intuitively, the least favorable prior should make the alternative hypothesis hard to distinguish. Hence, a rule of thumb consists in concentrating Λ on the boundary between H_1 and H_0 (i.e. the line $\{\sigma = \sigma_0\}$). Thus Λ will be a probability distribution over $\theta \in \mathbb{R}$ for the fixed $\sigma = \sigma_0$.

Another useful observation is that, given any test function $\phi(x)$ and a sufficient statistic T, there exists a test function η that has the same power as ϕ but depends on x only through T:

$$\eta(T(x)) = \mathbb{E}[\phi(x)|T(x)].$$

Hence, we can restrict our attention to the sufficient statistics (Y, U), where $Y = \bar{X}$ and $U = \sum_{i=1}^n (X_i - \bar{X})^2$. We know that $Y \sim \mathcal{N}(\theta, \sigma^2/n)$, $U \sim \sigma^2\chi^2_{n-1}$, and Y is independent of U by Basu's theorem.

Thus, for Λ supported on $\sigma = \sigma_0$, we obtain the joint density of (Y, U) under H_Λ as

$$c_0 u^{\frac{n-3}{2}} \exp\left(-\frac{u}{2\sigma_0^2}\right) \int \exp\left(-\frac{n}{2\sigma_0^2}(y - \theta)^2\right) d\Lambda(\theta)$$

and the joint density under alternative hypothesis (θ_1, σ_1) as

$$c_1 u^{\frac{n-3}{2}} \exp\left(-\frac{u}{2\sigma_1^2}\right) \exp\left(-\frac{n}{2\sigma_1^2}(y - \theta_1)^2\right).$$

From the above observations, we see that the choice of Λ only affects the distribution of Y. To achieve minimal maximum power against the alternative (i.e., to be least favorable), we need to choose Λ such that the two distributions become as close as possible. Under the alternative hypothesis, $Y \sim \mathcal{N}\left(\theta_1, \frac{\sigma_1^2}{n}\right)$. Under H_Λ, the distribution of Y is in a convolution form, i.e., $Y = Z + \Theta$ for $Z \sim \mathcal{N}\left(0, \frac{\sigma_0^2}{n}\right)$, $\Theta \sim \Lambda$, where Z and Θ are independent. Hence, if we choose $\Theta \sim \mathcal{N}\left(\theta_1, \frac{\sigma_1^2 - \sigma_0^2}{n}\right)$, Y will have the same distribution under the null and the alternative, which is $\mathcal{N}\left(\theta_1, \frac{\sigma_1^2}{n}\right)$. Under this choice of prior, the LRT rejects for large values of $\exp\left(-\frac{u}{2\sigma_1^2} + \frac{u}{2\sigma_0^2}\right)$, i.e., it rejects for large values of u (since $\sigma_1 > \sigma_0$). So the MP test rejects H_Λ if $\sum_{i=1}^n (X_i - \bar{X})^2$ lies above some threshold determined by the size constraint. In particular, it rejects if $\sum_{i=1}^n (X_i - \bar{X})^2 > \sigma_0^2 C_{n-1,1-\alpha}$, where $C_{n-1,1-\alpha}$ is the $(1-\alpha)^{th}$ quantile of χ^2_{n-1}.

3. Next we check if the MP test is level-α for the composite null. For any (θ, σ) with $\sigma \leq \sigma_0$, the probability of rejection is:

$$\mathbb{P}_{\theta,\sigma}\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} > \frac{\sigma_0^2 \, C_{n-1,1-\alpha}}{\sigma^2}\right) = \mathbb{P}\left(\chi^2_{n-1} > \frac{\sigma_0^2}{\sigma^2}C_{n-1,1-\alpha}\right) \leq \alpha,$$

while equality holds if $\sigma = \sigma_0$. Hence, it follows from Theorem 1 that our test is MP for testing the original null H_0 vs. $\mathcal{N}(\theta_1, \sigma_1)$.

4. Finally, the MP level-α test for testing the composite null H_0 vs. an arbitrarily chosen (θ_1, σ_1) does not depend on the choice of (θ_1, σ_1). Hence it is UMP for testing the original composite null vs. the composite alternative.

Example 2 (Nonparametric Quality Checking)**.** Identical light bulbs have lifetime $X_1, \ldots, X_n$ with an arbitrary distribution $\mathbb{P}$ over $\mathbb{R}$. Let u be a fixed threshold for a satisfactory lifetime and $\mathbb{P}(X \leq u)$ be the probability of a given light bulb being unsatisfactory. Given the data of sample lifetimes we may be interested in testing whether the probability of having an unsatisfactory light bulb is too large:

$$H_0 : \mathbb{P}(X \leq u) \geq p_0 \ \text{ vs. } \ H_1 : \mathbb{P}(X \leq u) < p_0.$$

Here p_0 is a fixed quality parameter.

0. Before we start our search for the UMP test, let us reparametrize the distribution $\mathbb{P}$ as follows. Let $\mathbb{P}^-$ and $\mathbb{P}^+$ be the conditional distributions of $X|X \leq u$ and $X|X > u$ respectively, and let $p = \mathbb{P}(X \leq u)$. Then, $\mathbb{P}$ has a one-to-one correspondence with $(\mathbb{P}^-, \mathbb{P}^+, p)$. For any fixed $\mathbb{P}$, let p^- and p^+ be the conditional densities of $\mathbb{P}^-$ and $\mathbb{P}^+$ with respect to some measure μ (existence of the densities and base measure can be justified, e.g. by Radon-Nikodym theorem is necessary). The joint density of $X_1, \ldots, X_n$ at values $x_1, \ldots, x_n$ when $x_{i_1}, \ldots, x_{i_m} \leq u < x_{j_1}, \ldots, x_{j_{n-m}}$ is then given by

$$p^m\left(\prod_{j=1}^m p^-(x_{i_j})\right)(1-p)^{n-m}\left(\prod_{k=1}^{n-m} p^+(x_{j_k})\right).$$

1. As before, we fix a simple alternative $(\mathbb{P}^-, \mathbb{P}^+, p_1)$ where $p_1 < p_0$.

2. We next choose a proper prior. We guess that Λ mostly concentrates on the boundary point (p^+, p^-, p_0). If so, for testing H_Λ vs. the simple alternative, the LRT rejects for large values of

$$\frac{\binom{n}{m}p_1^m(1-p_1)^{n-m}}{\binom{n}{m}p_0^m(1-p_0)^{n-m}},$$

which is equivalent to testing $Bin(n, p_0)$ vs. $Bin(n, p_1)$. Thus, the MP test, which rejects for small values of $m = \#\{i : X_i \leq u\}^1$, is given by

$$\phi_\Lambda(x) = \begin{cases} 1, & \text{if } m < k \\ \gamma, & \text{if } m = k \\ 0, & \text{if } m > k, \end{cases}$$

where k and γ are both determined by the level constraint $\mathbb{E}_{p_0}\phi_\Lambda(x) = \alpha$.

3. Now we check if ϕ_Λ is level-α for our composite null H_0. Note that the power function of ϕ_Λ depends on $\mathbb{P}$ only through $p = \mathbb{P}(X \leq u)$. Given that this family has MLR in m, the power function would be monotone. So for any $p > p_0$, the rejection probability under the null is still smaller than α. Hence, ϕ_Λ is the MP test for testing the composite null H_0 against the simple alternative $H_1 : (\mathbb{P}^-, \mathbb{P}^+, p_1)$.

4. Finally, ϕ_Λ has no dependence on the choice of alternative hypothesis. Therefore, ϕ_Λ is UMP for testing the composite null H_0 against the composite alternative H_1.

Exercise 13 (#6.18). Let $(X_1, ..., X_n)$ be a random sample from the uniform distribution on $(\theta, \theta + 1)$, $\theta \in \mathcal{R}$. Suppose that $n \geq 2$.
(i) Show that a UMP test of size $\alpha \in (0, 1)$ for testing $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$ is of the form

$$T_*(X_{(1)}, X_{(n)}) = \begin{cases} 0 & X_{(1)} < 1 - \alpha^{1/n}, \ X_{(n)} < 1 \\ 1 & \text{otherwise,} \end{cases}$$

where $X_{(j)}$ is the jth order statistic.
(ii) Does the family of all densities of $(X_{(1)}, X_{(n)})$ have monotone likelihood

Solution A. (i) The Lebesgue density of $(X_{(1)}, X_{(n)})$ is

$$f_\theta(x, y) = n(n-1)(y-x)^{n-2}I_{(\theta,y)}(x)I_{(x,\theta+1)}(y).$$

A direct calculation of $\beta_{T_*}(\theta) = \int T_*(x, y)f_\theta(x, y)dxdy$, the power function of T_*, leads to

$$\beta_{T_*}(\theta) = \begin{cases} 0 & \theta < -\alpha^{1/n} \\ (\theta + \alpha^{1/n})^n & -\alpha^{1/n} \leq \theta \leq 0 \\ 1 + \alpha - (1-\theta)^n & 0 < \theta \leq 1 - \alpha^{1/n} \\ 1 & \theta > 1 - \alpha^{1/n}. \end{cases}$$

For any $\theta_1 \in (0, 1 - \alpha^{1/n}]$, by the Neyman-Pearson Lemma, the UMP test T of size α for testing $H_0 : \theta = 0$ versus $H_1 : \theta = \theta_1$ is

$$T = \begin{cases} 1 & X_{(n)} > 1 \\ \alpha/(1-\theta_1)^n & \theta_1 < X_{(1)} < X_{(n)} < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The power of T at θ_1 is computed as

$$\beta_T(\theta_1) = 1 - (1-\theta_1)^n + \alpha,$$

which agrees with the power of T_* at θ_1. When $\theta > 1 - \alpha^{1/n}$, T_* has power 1. Therefore T_* is a UMP test of size α for testing $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$.
(ii) The answer is no. Suppose that the family of densities of $(X_{(1)}, X_{(n)})$ has monotone likelihood ratio. By the theory of UMP test (e.g., Theorem 6.2 in Shao, 2003), there exists a UMP test T_0 of size $\alpha \in (0, \frac{1}{2})$ for testing $H_0 : \theta \leq 0$ versus $H_1 : \theta > 0$ and T_0 has the property that, for $\theta_1 \in (0, 1-\alpha^{1/n})$, T_0 is UMP of size $\alpha_0 = 1+\alpha-(1-\theta_1)^n$ for testing $H_0 : \theta \leq \theta_1$ versus $H_1 : \theta > \theta_1$. Using the transformation $X_i - \theta_1$ and the result in (i), the test

$$T_{\theta_1}(X_{(1)}, X_{(n)}) = \begin{cases} 0 & X_{(1)} < 1 + \theta_1 - \alpha_0^{1/n}, X_{(n)} < 1 + \theta_1 \\ 1 & \text{otherwise} \end{cases}$$

is a UMP test of size α_0 for testing $H_0 : \theta \leq \theta_1$ versus $H_1 : \theta > \theta_1$. At $\theta = \theta_2 \in (\theta_1, 1-\alpha^{1/n}]$, it follows from part (i) of the solution that the power of T_0 is $1+\alpha-(1-\theta_2)^n$ and the power of T_{θ_1} is $1+\alpha_0-[1-(\theta_2-\theta_1)]^n$. Since both T_0 and T_{θ_1} are UMP tests, $1+\alpha-(1-\theta_2)^n = 1+\alpha_0-[1-(\theta_2-\theta_1)]^n$. Because $\alpha_0 = 1 + \alpha - (1-\theta_1)^n$, this means that

$$1 = (1-\theta_1)^n - (1-\theta_2)^n + [1 - (\theta_2 - \theta_1)]^n$$

holds for all $0 < \theta_1 < \theta_2 \leq 1-\alpha^{1/n}$, which is impossible. This contradiction proves that the family of all densities of $(X_{(1)}, X_{(n)})$ does not have monotone likelihood ratio.

Problem 3.16 Based on X with distribution indexed by $\theta \in \Omega$, the problem is to test $\theta \in \omega$ versus $\theta \in \omega'$. Suppose there exists a test ϕ such that $E_\theta[\phi(X)] \leq \beta$ for all θ in ω, where $\beta < \alpha$. Show there exists a level α test $\phi^*(X)$ such that

$$E_\theta[\phi(X)] \leq E_\theta[\phi^*(X)] \,,$$

for all θ in ω' and this inequality is strict if $E_\theta[\phi(X)] < 1$.

if $E_\theta[\phi(X)] < 1$
$\phi^*(X) = \min\{\phi(X) + \alpha - \beta, 1\}$
$\forall\theta\in\omega$, $E_\theta\phi^*(X) \leq E_\theta[\phi(X) + \alpha - \beta] \leq \beta + \alpha - \beta = \alpha$
$\forall\theta\in\omega'$ $\because E_\theta\phi(X) < 1 \Rightarrow P_\theta(\phi(X) < 1) > 0$
$\exists\varepsilon > 0$ st. $P_\theta(\phi(X) < 1 - \varepsilon) > 0$ and $\varepsilon < \alpha - \beta$

$$\begin{aligned}
E_\theta\phi^*(X) &= E_\theta\phi^*(X)1_{\{\phi(X) < 1-\varepsilon\}} + E_\theta\phi^*(X)1_{\{1 \geq \phi(X) \geq 1-\varepsilon\}} \\
&\geq E_\theta\phi(X)1_{\{\phi(X) < 1-\varepsilon\}} + \varepsilon E_\theta 1_{\{\phi(X) < 1-\varepsilon\}} + E_\theta\phi(X)1_{\{1 \geq \phi(X) \geq 1-\varepsilon\}}
\end{aligned}$$

$\phi(X) + \alpha - \beta > \phi(X) + \varepsilon$
$1 > \phi(X)$

$\phi(X) + \alpha - \beta > \phi(X)$
$1 > \phi(X)$

$= E_\theta(\phi(X)) + \varepsilon P[\phi(X) < 1 - \varepsilon] > E_\theta\phi(X)$

Suppose we want to test H_0 vs H_1 at level α. Let ϕ_α test at level α. Assume that the rejection regions an nested, i.e. $\alpha_1 < \alpha_2 \Rightarrow S_{\alpha_1} \subseteq S_{\alpha_2}$ where $S_\alpha = \{x : \phi_\alpha(x) = 1\}$.

Definition 9.1 $\hat{p}(x) = \inf\{u : x \in \mathrm{S_u}\}$.

Intuitively, given the p-value, you can construct a level α-test by rejecting H_0 if $\hat{p}(x) < \alpha$ and accepting H_0 if $\hat{p}(x) > \alpha$, e.g.

$$\begin{aligned} S_\alpha &= \left\{X_i \sum X_i > Z_{1-\alpha}\sqrt{n}\right\} \\ &= \left\{X_: 1 - \Phi\left(\frac{\Sigma X_i}{\sqrt{n}}\right) < \alpha\right\} \\ &\Rightarrow \hat{p}(x) = 1 - \Phi\left(\frac{\Sigma x_i}{\sqrt{h}}\right) \end{aligned}$$

under $H_0 : g = 0, \hat{p}(x) \sim U(0,1)$

$$P(\hat{p}(x) \leq u) = u$$

Lemma 9.2 *Suppose $X \sim P_\theta$ for some $\theta \in \Theta$. We want to test $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ at level α. Let $\{\phi_\alpha\}_{\alpha \in (0,1)}$ be a collection of nested level α tests. Then*
(i) $P_\theta(\hat{p}(x) \leq u) \leq u = P(u(0,1) \leq u), \forall u \in (0,1), \theta \in \Theta_0$
(ii) If $\exists g_0 \in \Theta_0$, such that $P_{\theta_0}(X \in S_\alpha) = \alpha$ for $\forall \alpha$, then $P_{\Theta_0}(\hat{p}(x) \leq u) = u$.

Definition 9.3 ***(Confidence interval)*** *Let $X \sim P_\theta$ for some $\theta \in \Theta$. For every $X \in X$, Let $S(X)$ be a subset of Θ. We say the collection of sets $\{S(X), X \in \mathcal{X}\}$ is an $(1-\alpha)$ confidence region if $P_\theta(\theta \in S(x)) \geq 1 - \alpha, \forall \theta \in \Theta$.*

9.2 Asymptotic Optimality

Let $\{X_1, \cdots, X_n\}$ be i.i.d. from $\{P_\theta, 0 \in \Theta\}$ with pdf w.r.t. some σ-finite measure. Suppose we want to estimate $g(\theta)$, and a candidate estimator is $\delta_n(x_1, \ldots, x_n)$.

Definition 9.4 *We say $\delta_n(x)$ is consistent for $g(\theta)$, if $\delta_n(x) \xrightarrow{p} g(\theta), \quad \forall \theta \in \theta$,*
i.e. $\forall \theta \in \Theta, \forall \varepsilon > 0$, we have $P_\theta(|\delta_n(X) - g(\theta)| > \epsilon) \to 0$.

Remark: If $X_1 \ldots X_n \overset{\text{iid}}{\sim}$ F:
(i)Assume $E_F|X| < \infty$, then $\frac{1}{n}\sum x_i \xrightarrow{p} E_F(X)$ (WLLN).
(ii) Assume $E_F X^2 < \infty$, then $W_n = \frac{\sum X_i - nE_p(X)}{\sqrt{n\,\mathrm{var_p}(X)}} \xrightarrow{d} \mathcal{N}(0,1)$(CLT),

$$\Leftrightarrow \lim_{n\to\infty} P(W_n \leq t) = \Phi(t), \quad \forall t \in R$$

(Example) $X_1 \ldots X_n \overset{\text{iid}}{\sim}$ Bernoulli (θ), if $\theta \in Q$; $X_1 \ldots X_n \overset{\text{iid}}{\sim}$ Bernoulli $(1-\theta)$. if $\theta \notin Q$; then there is no consistent estimator of $X_1 \ldots X_n$.

Definition 9.5 *Let $L(\theta|x) = \prod_{i=1}^n p_\theta(x_i)$ be the likelihood function. If there exists a unique θ_n which is the global maximizer of $\theta \to L(\theta/x)$(or $\theta \to l(\theta \mid x) = \log L(\theta \mid x)$). Define $\hat{\theta}_n$ as the MLE of θ.*

(Example) $X_1 \ldots X_n \overset{\text{iid}}{\sim}$ Bernoulli (θ). $\log(\theta|x) = \sum_{i=1}^n x_i(\log\theta) + (n - \sum_{i=1}^n x_i)\log(1-\theta)$,
$\Rightarrow \hat{\theta}_n = \arg\max_\theta \ln(\theta \mid x) = \frac{\sum_{i=1}^n x_i}{n} = \sum_{i=1}^n x_i/n$.
(i) $\bar{X}_n \xrightarrow{p} \theta, \quad \forall \theta \in (0,1)$, (consistency)
(ii) $\sqrt{n}(\overline{x_n} - \theta) \xrightarrow{d} N\left(0.\frac{1}{\theta(1-\theta)}\right)$, (CLT).

Theorem 9.6 *Suppose $X_1 \ldots X_n \overset{iid}{\sim} P_\theta$ for some $\theta \in \Theta$ with pdf $P_\theta(\cdot)$*
Assume A0: $P_{\theta_1} \neq P_{\theta_2}$. where $\theta_1 \neq \theta_2$ [identifiability]; A1: $\{P_\theta(\cdot), \theta \in \Theta\}$ has common support.
Then we have: $P_{\theta_0}(\log(\hat{\theta n} \mid x) > \log(\theta \mid x)) \to 1$ as $n \to \infty, \quad \forall \theta \neq \theta_0$.

Proof: Let $T_n = \frac{1}{n}\sum_{i=1}^n \log\frac{p_\theta(x_i)}{p_\theta(x_i)}$, then $T_n \xrightarrow{p} E_{\theta_0}\log\frac{p_\theta(x_1)}{p_{\theta_0}(x_2)}$. Now $E_\theta \log\frac{P_\theta(X_1)}{P_{\theta_0}(X_1)} = \int \log\frac{p_0(x)}{p_\theta(x)} p_\theta(x) d\mu(x) = -D(\theta_0||\theta) < 0$ for $\theta \neq \theta_0$. Hence, $P_{\theta_0}(T_n \neq 0) \to 1$, but $T_n < 0 \Leftrightarrow \frac{1}{n}\sum_{i=1}^n \log\frac{P_\theta(X_i)}{P_\theta(X_i)} < 0 \Leftrightarrow \log\prod_{i=1}^n P_\theta(X_i) < \log\prod_{i=1}^n P_{\theta_0}(X_i) \Leftrightarrow \ell_n(\theta|x) < \ell_n(\theta|x)$. ■

Corollary 9.7 *Suppose A0 and A1 hold, it Θ is finite, then the MLE $\hat{\theta_n}$ exists with high probability and $P_{\theta_0}(\hat{\theta_n} = 0) \to 1, \quad (n \to \infty)$.*

Suppose $A0$ and $A1$ hold. Suppose that $\Theta \subseteq \mathbb{R}$ and θ_0 is an interior point of Θ. If $\theta \mapsto p_0(x)$ is differentiable and the deviates is continuous. there exist a sequence of roots $\hat{e}_n$ of the likelihood equation $\frac{\partial}{d\theta}\mathrm{In}(\theta/x) = 0$. where is consistent for θ_0.

Let $A_n = \{x : \ln(\theta_0 \mid x) > \max_{j \leq k}\ln(\theta_j \mid x)\}$. If $X \in A_n$, then $\hat{\theta_n}(x) = \theta_0$ and $P_{\theta_0}(A_n) \to 1$.

Theorem 9.8 *Suppose A0 and A1 hold. Suppose further that A2: $\Theta \subset \mathbb{R}$ and θ_0 is an interior point of Θ. If $\theta \mapsto P_\theta()$ is differentiable and the derivative is continuous, then there exists a sequence of roots $\hat{\theta}_n$ of the score function $\ell_n'(\theta)\partial\ell_n(\theta\,|)/\partial\theta = 0$, which is consistent for θ_0.*

Proof: Let $\delta > 0$ be small enough such that $[\theta_0 - \delta, \theta_0 + \delta] \subset \Theta$. It follows that

$$P_{\theta_0}(\ell_n(\theta_0\,|) > \ell_n(\theta_0 \pm \delta\,|)) \to 1$$

as $n \to \infty$. Now, the function $\theta \mapsto \ell_n(\theta\,|)$ is a continuous function on the compact set $[\theta_0 - \delta, \theta_0 + \delta]$. There exists a global maximiser $\tilde{\theta}_n(\delta)$. But $\tilde{\theta}_n(\delta)$ cannot be $\theta_0 \pm \delta$ as θ_0 is better, which implies that $\ell_n'(\tilde{\theta}_n(\delta)) = 0$.

Let $\hat{\theta}_n(\delta)$ denote the closest root of $\ell_n'(\theta) = 0$ to θ_0. Fix $\delta > 0$, we need to show that $P_{\theta_0}(|\hat{\theta}_n - \theta_0| < \delta) \to 1$ as $n \to \infty$. Observe that $|\hat{\theta}_n - \theta_0| \leq |\tilde{\theta}_n(\delta) - \theta_0|$ as $\hat{\theta}_n$ is the closet root. It follows that

$$P_{\theta_0}(|\hat{\theta}_n - \theta_0| < \delta) \geq P_{\theta_0}(|\tilde{\theta}_n(\delta) - \theta_0| < \delta) \geq P_{\theta_0}(\ell_n(\theta_0) > \ell_n(\theta_0 \pm \delta)) \to 1.$$

It remains to prove that there exists a closest root, i.e. $\exists\hat{\theta}$ such that $f(\hat{\theta}) = 0, |\hat{\theta} - \theta_0| = \inf_{\tilde{\theta}: f(\tilde{\theta})=0}|\tilde{\theta} - \theta_0|$, assuming that $\{\tilde{\theta} : f(\tilde{\theta}) = 0\}$ is non-empty, and $f(\cdot)$ is a continuous function on $\mathbb{R}$. To see this, let $\alpha = \inf_{\tilde{\theta}: f(\tilde{\theta})=0}|\tilde{\theta} - \theta_0|$. For all $k \geq 1$, there exists $\tilde{\theta}_k$ such that

$$f(\tilde{\theta}_k) = 0 \quad \text{and} \quad |\tilde{\theta}_k - \theta_0| \leq \alpha + k^{-1} \leq \alpha + 1. \tag{9.1}$$

Note also that $\tilde{\theta}_k \in [\theta_0 - \alpha - 1, \theta_0 + \alpha + 1]$. By going to a subsequence, as $k \to \infty$, $\tilde{\theta}_k \to \hat{\theta}$, say. But $|\hat{\theta} - \theta_0| = \alpha$ by taking the limit on (9.1) and the fact that $f(\hat{\theta}) = 0$ since $f(\cdot)$ is continuous. ■

Corollary 9.9 *If A0–A2 hold, assume further that $\theta \mapsto P_\theta()$ is differentiable, and the score function $\ell_n'(\theta) = 0$ has a unique root $\hat{\theta}_n$, then $\hat{\theta}_n \xrightarrow{P} \theta_0$ (follows from the previous theorem), and $\hat{\theta}_n$ is the MLE with probability tending to 1.*

Proof: It follows from the previous proof that $\hat{\theta}_n$ is a local maximum (with high probability). If $\hat{\theta}_n$ is not the unique global mimimiser of $\theta \mapsto \ell_n(\theta)$, then there exists $\tilde{\theta}_n$ such that $\ell_n(\tilde{\theta}_n) \geq \ell_n(\hat{\theta}_n), \tilde{\theta} \neq \hat{\theta}_n$. Then there exists $\breve{\theta}$ such that $\ell_n(\breve{\theta}) = \ell_n(\hat{\theta}_n), \breve{\theta} \neq \hat{\theta}$, as $\theta \mapsto \ell_n(\theta)$ is continuous. It implies that there exists $\epsilon_n \neq \hat{\theta}_n$ such that $\ell_n'(\epsilon_n) = 0$ [see Rolle's Theorem], which is a contradiction. ■