

Advanced Statistical Inference

STAT 561 - Advanced Statistical Inference

Dr. Bing Li



Copyright © 2013 John Smith

PUBLISHED BY PUBLISHER

BOOK-WEBSITE.COM

Licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License (the “License”). You may not use this file except in compliance with the License. You may obtain a copy of the License at <http://creativecommons.org/licenses/by-nc/3.0>. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

First printing, March 2013

Contents

I	Part One	
1	Basic Ideas in Bayesian Analysis	7
1.1	Frequentist & Bayesian Settings	14
1.2	Prior Posterior & Likelihood	15
1.3	Conditional Independence and Frequentist/Bayesian Sufficiency	15
1.4	Equivalence of Frequentist & Bayesian Sufficiency	18
1.5	Conjugate Priors	21
1.5.1	Introduction	21
1.5.2	Exponential Family	21
1.5.3	Convex Hull of Conjugate Family	23
1.6	Two-Parameter, Normal Family	24
1.7	Noninformative Prior	27
1.7.1	General Concept	27
1.7.2	Invariant Prior	28
1.7.3	Jeffreys' Prior	32
1.8	Statistical Decision Theory	33
2	Estimation & Inference	39
2.1	Estimation	39
2.2	Bayes Rule and Unbiasedness	43
2.3	Error Assessment in Bayesian Setting	44
2.4	Credible Set (or Interval)	45
2.5	Hypothesis Test	47

2.6	Classificaiton	52
2.7	Stein's Estimate	53
2.8	Empirical Bayes	56
3	Testing Hypothesis for Simple-VS-Simple	61
3.1	Basic Concepts	61
3.2	Neyman-Pearson Lemma	62
3.3	Uniformly Most Powerful Test for One-Sided Hypothesis	66
3.3.1	General Definition	66
3.3.2	Monotone Likelihood Ratio Assumption	67
3.3.3	General Form of UMP-Test for One-Sided Hypothesis	67
3.3.4	Properites of UMP- α Test for One Sided Hypotheses	68
3.4	Uniformly Most Powerful Unbiased Test - 2 Sided Hypothesis	70
3.4.1	Uniformly Most Powerful Unbiased Test (UMPU)	71
3.4.2	Exponential Family	72
3.4.3	Generlized Neyman Pearson Lemma	73
3.4.4	UMP-Test for Test III	74
3.4.5	UMPU- α Test for II, III	75
4	Testing Hypotheses in the Presense of Nuisance Pamameters ...	79
4.1	Unbiased and Similar Tests	80
4.2	Sufficient and Complete	82
4.3	UMPU test in the presence of Nusience Parameter	84
4.4	Using Basu's Theorem to construct UMPU in presence of nuisance paramet- ters	84

Part One

1	Basic Ideas in Bayesian Analysis	7
1.1	Frequentist & Bayesian Settings	
1.2	Prior Posterior & Likelihood	
1.3	Conditional Independence and Frequentist/Bayesian Sufficiency	
1.4	Equivalence of Frequentist & Bayesian Sufficiency	
1.5	Conjugate Priors	
1.6	Two-Parameter, Normal Family	
1.7	Noninformative Prior	
1.8	Statistical Decision Theory	
2	Estimation & Inference	39
2.1	Estimation	
2.2	Bayes Rule and Unbiasedness	
2.3	Error Assessment in Bayesian Setting	
2.4	Credible Set (or Interval)	
2.5	Hypothesis Test	
2.6	Classification	
2.7	Stein's Estimate	
2.8	Empirical Bayes	
3	Testing Hypothesis for Simple-VS-Simple	61
3.1	Basic Concepts	
3.2	Neyman-Pearson Lemma	
3.3	Uniformly Most Powerful Test for One-Sided Hypothesis	
3.4	Uniformly Most Powerful Unbiased Test - 2 Sided Hypothesis	
4	Testing Hypotheses in the Presence of Nuisance Parameters	79
4.1	Unbiased and Similar Tests	
4.2	Sufficient and Complete	
4.3	UMPU test in the presence of Nuisance Parameter	
4.4	Using Basu's Theorem to construct UMPU in presence of nuisance parameters	

1. Basic Ideas in Bayesian Analysis

Mathematical Preparation

Monday January 9

1. Product σ -Field

$(\Omega_1, \mathcal{F}_1, \mu_1), (\Omega_2, \mathcal{F}_2, \mu_2)$ are two measure spaces. The goal is to construct a σ -field on $\Omega_1 \times \Omega_2$.

Let $\mathcal{A} = \{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$.

The σ -field generated \mathcal{A} is called the product σ -field, written as $\mathcal{F}_1 \times \mathcal{F}_2$, that is $\sigma(\mathcal{A})$. This is NOT a cartesian product, which would be $\{(A, B) : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$.

2. Product Measure

Let $E \in \mathcal{F}_1 \times \mathcal{F}_2$. Let $E_2(\omega_1) = \{\omega_2 : (\omega_1, \omega_2) \in E\}$ and similarly, $E_1(\omega_2) = \{\omega_1 : (\omega_1, \omega_2) \in E\}$.

It is true (in Billingsly) that

Theorem 1.0.1 — Number Unknown. If $E \in \mathcal{F}_1 \times \mathcal{F}_2$ then $E_1(\omega_2) \in \mathcal{F}_1$ for all $\omega_2 \in \Omega_2$. Similarly, $E_2(\omega_1) \in \mathcal{F}_2$ for all $\omega_1 \in \Omega_1$.

If $f : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ measurable $\mathcal{F}_1 \times \mathcal{F}_2 \setminus \mathcal{R}$. Then for each $\omega_1 \in \Omega_1$,

$$f(\omega_1, \cdot) \in \mathcal{R} \text{ for each } \omega_1 \in \Omega_1.$$

$$f(\cdot, \omega_2) \otimes \mathcal{F}_1 \setminus \mathcal{R}$$

Now, for each $E \in \mathcal{F}_1 \times \mathcal{F}_2$ consider

$$f_{1,E} : \Omega_1 \rightarrow \mathcal{R}, \omega_1 \mapsto \mu_2(E_2, (\omega_2))$$

It can be shown that $f_{1,E}$ is uniformly measurable $\mathcal{F}_1 \setminus \mathcal{R}$ for all E .

Proof. Outline.

- Show that if $\mathcal{L} = \{E : f_{1,E} \otimes \mathcal{F}_1 \setminus \mathcal{R}\}$ then \mathcal{L} is a λ -system.
 - Let $\mathcal{P} = \{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$ then it is a π -system.
- Furthermore, if $E = A \times B$,

$$E_2(\omega_1) = \begin{cases} B & \omega_1 \in A \\ \emptyset & \omega_1 \notin A \end{cases}$$

$$\text{So, } \mu_2(E_2(\omega_1)) = \begin{cases} \mu_2(B) & \omega_1 \in A \\ 0 & \omega_1 \notin A \end{cases} = I_A(\omega_1)\mu(B) = f_{1,E}$$

So, $f_{1,E} \otimes \mathcal{F}_1$.

Thus $\mathcal{P} \subseteq \mathcal{L}$.

- By $\pi - \lambda$ Theorem, $\mathcal{F}_1 \times \mathcal{F}_2 \subseteq \mathcal{L}$.

Similarly, $f_{2,E} \otimes \mathcal{F}_2 \setminus \mathcal{R}$.

We can now define two set functions,

$$\pi'(E) = \int f_{1,E} d\mu_1$$

$$\pi''(E) = \int f_{2,E} d\mu_2$$

Again using $\pi - \lambda$ Theorem, it can be shown that, π', π'' are both measure and if μ_1, μ_2 are σ -finite, then

$$\pi' = \pi'' \text{ on } \mathcal{F}_1 \times \mathcal{F}_2$$

Note that here, \mathcal{P} equals \mathcal{A} used at beginning of notes.

We did not have a measure in $\mathcal{F}_1 \times \mathcal{F}_2$. Now we have π', π'' both measures on $\mathcal{F}_1 \times \mathcal{F}_2$, they are the same. We call this measure the product measure, written as $\mu_1 \times \mu_2$.

Note that $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mu_1 \times \mu_2)$ is called product measure space. ■

3. Tonelli's Theorem

$(\Omega_1, \mathcal{F}_1, \mu_1), (\Omega_2, \mathcal{F}_2, \mu_2)$ are two σ -finite measure spaces.

$(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mu_1 \times \mu_2)$ is the product measure space.

Suppose we have $f : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R} \otimes \mathcal{F}_1 \times \mathcal{F}_2 \setminus \mathcal{R}$. Where $f \geq 0$ and

$$\int f d(\mu_1 \times \mu_2) = \int \left[\int (f(\cdot, \omega_2) d\mu_1) \right] d\mu_2$$

4. Fubini's Theorem

The conclusion of Tonelli's Theorem still holds if f is NOT nonnegative, but if f is integrable μ_2 . (integrable - integral of absolute value of function is finite)

Wednesday January 11

5. Conditional Probability

This is a special application of Radon- Nikodgm Theorem. We know that

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

We may define $P(A|\mathcal{G})$ when $\mathcal{G} \subseteq \mathcal{F}$ as sub- σ -field. We defined this intuitively in elementary probability course (definition above), but we are not going to define it generally.

Now let $A \in \mathcal{F}$ and $\mathcal{G} \subset \mathcal{F}$ be a σ -field. Consider the set function

$$\nu : \mathcal{G} \rightarrow \mathbb{R}, G \mapsto P(AG)$$

It can be easily shown that ν is a measure on \mathcal{G} . Consider another set function,

$$\mu : \mathcal{G} \rightarrow \mathbb{R}, G \mapsto P(G)$$

So μ is nothing but P restricted on \mathcal{G} .

It's easy to show that $\nu \ll \mu$.

$$\mu(G) = 0 \Rightarrow P(G) = 0 \Rightarrow P(AG) = 0 \Rightarrow \nu(G) = 0$$

By Radon-Nikodgm Theorem, there exists a δ such that

$$\nu(G) = \int_G \delta d\mu \quad \forall G \in \mathcal{G}$$

δ is called R-N Derivative, written as

$$\delta = \frac{d\nu}{d\mu}$$

and is similar in to $\frac{P(AG)}{P(G)}$, but it's more general.

δ is called the conditional probability of A given \mathcal{G} . To distinguish it from $P(A|B)$, where B is a set, we use $P(A|\mathcal{G})$, where \mathcal{G} is a σ -field. By construction,

- (a) δ is measurable \mathcal{G}
- (b) $\int_G \delta d\mu = P(AG) \quad \forall G \in \mathcal{G}$

Note that, by RNT, δ is unique with probability 1. Any δ' satisfying (a) and (b) has $\delta' = \delta a.e.P$. So, we say that δ is a version of conditional probability.

So, δ is a version of $P(A|\mathcal{G})$ if and only if (a) and (b) are satisfied. We may define $P(A|\mathcal{G})$ either by RNT or (a) and (b).

Properties of Conditional Probability

It behaves like probability, but since it is a function, unique up to a.e. P, these properties have to be qualified by a.s. P.

$$(a) P(\emptyset|\mathcal{G}) = 0, P(\Omega|\mathcal{G}) = 1 \text{ a.s. P}$$

$$(b) 0 \leq P(A|\mathcal{G}) \leq 1 \text{ a.s. P}$$

$$(c) \text{ If } A_1, A_2, \dots \text{ are disjoint members of } \mathcal{F} \text{ then } P(\bigcup_n A_n|\mathcal{G}) = \sum_n P(A_n|\mathcal{G}) \text{ a.s. P}$$

Let's consider the special case where \mathcal{G} is a σ -field generated by some random element, T (i.e. $\mathcal{G} = \sigma(T)$). More specifically, for some measurable space $(\Omega_T, \mathcal{F}_T)$ where

$$T : \Omega \rightarrow \Omega_T \text{ in } \mathcal{F} \setminus \mathcal{F}_T \quad \mathcal{G} = T^{-1}(\mathcal{F}_T)$$

Here, we write

$$\begin{aligned} P(A|\mathcal{G}) &= P(A|\sigma(T)) \\ &= P(A|T^{-1}(\mathcal{F}_T)) \\ &= P(A|T) \end{aligned}$$

The following theorem makes checking that something is a conditional probability easier. In principle, we have to check $\int_G \delta dp = P(AG) \quad \forall G \in \mathcal{G}$.

Theorem 1.0.2 — 33.1 in Billingsly. Let \mathcal{P} be a π -system generating \mathcal{G} and suppose that Ω is a countable union of sets in \mathcal{P} . An integrable function, f , is a version of $P(A|\mathcal{G})$ if

$$(a) f \text{ is measurable } \mathcal{G}$$

$$(b) \int_G f dp = P(AG) \quad \forall G \in \mathcal{P}$$

6. Conditional Distribution

Let there be probability space (Ω, \mathcal{F}, P) , measurable space $(\Omega_X, \mathcal{F}_X)$, and a random element, $X : \Omega \rightarrow \Omega_X \text{ in } \mathcal{F} \setminus \mathcal{F}_X$. Also, let $\mathcal{G} \subseteq \mathcal{F}$ be a sub σ -field.

We are going to define conditional distribution of X given G. Under very mild conditions there is a function

$$f : \mathcal{F}_X \times \Omega \rightarrow \mathbb{R}$$

such that for each $A \in \mathcal{F}_X$, $f(A, \cdot)$ is a version of

$$P(X \in A|\mathcal{G}) = P(X^{-1}(A)|\mathcal{G})$$

and, for each $\omega \in \Omega$, $f(\cdot, \omega)$ is a probability measure on $(\Omega_X, \mathcal{F}_X)$.

The only condition for this existence is $(\Omega_X, \mathcal{F}_X)$ must be a Borel Space, that is \mathcal{F}_X is Borel σ -field. This should always be the case for our purposes.

7. Conditional Expectation

Let us have the same probability space, measurable space, random element, and sub σ -field as defined before, but here with \mathbb{R} .

We want to define conditional expectation of X given \mathcal{G} .

First, assume $X \geq 0$. Consider a set function,

$$\nu : \mathcal{G} \rightarrow \mathbb{R}, G \mapsto \int_G X dP$$

It can be easily shown that ν is a measure.

Let μ again be $\mathcal{G} \rightarrow \mathbb{R}, G \mapsto P(G)$. Then $\nu \ll \mu$. By RNT, $\delta = \frac{d\nu}{d\mu}$ is well defined. This is defined to be conditional expectation of X given \mathcal{G} , written as

$$E(X|\mathcal{G})$$

Suppose $X \not\geq 0$, but integrable P . Recall that $X = X^+ - X^-$. Since $X^+, X^- \geq 0$, then both $E(X^+|\mathcal{G}), E(X^-|\mathcal{G})$ are defined by RNT. We define,

$$E(X|\mathcal{G}) = E(X^+|\mathcal{G}) - E(X^-|\mathcal{G})$$

Friday January 13

As in the case of $P(A|\mathcal{G})$, the equivalent conditions for $d : \Omega \rightarrow \mathbb{R}$ is a version of $E(X|\mathcal{G})$.

(a) δ measurable \mathcal{G}

(b) $\int_G \delta dP = \int_G X dP \quad \forall G \in \mathcal{G}$

INSERT PHOTO FROM BOARD - "Mesh"

The value of δ in each thick outlined cell is the average (with respect to P measure) of $X(\omega)$ over the subcells (thin outlined) in thick cells.

We see from this definition that if $A \in \mathcal{F}, X = I_A$ then the second condition becomes

$$\int_G \delta dP = \int_G I_A dP = P(A \cap G)$$

So, $E(I_A|\mathcal{G}) = P(A|\mathcal{G})$.

Properties of Conditional Expectations

Theorem 1.0.3 34.2 in Billingsly Suppose that X, Y, X_n are integrable P .

If $X = a$ a.e. P , then $E(X|\mathcal{G})$ a.s. P

(b) If $a, b \in \mathbb{R}$ then

$$E(aX + bY|\mathcal{G}) = a(E(X|\mathcal{G})) + b(E(Y|\mathcal{G})) \text{ a.s. } P$$

(c) If $X \leq Y$ a.s. P then

$$E(X|\mathcal{G}) \leq E(Y|\mathcal{G})$$

(d) $|E(X|\mathcal{G})| \leq E(|X|\mathcal{G})$ a.s. P (in fact this is true for all convex functions).

(e) If $X_n \rightarrow X$ a.s. P , $|X_n| \leq Y$, and Y integrable P , then

$$E(X_n|\mathcal{G}) \rightarrow E(X|\mathcal{G}) \text{ a.s. } P$$

Proof. Found in Billingsly. ■

Theorem 1.0.4 — 34.4 in Billingsly. If $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$ and X integrable P , then

$$E(E(X|\mathcal{G}_2)|\mathcal{G}_1) = E(X|\mathcal{G}_1)$$

This is called the Law of Iterative Conditional Expectation.

Theorem 1.0.5 — 34.3 in Billingsly. If X measurable \mathcal{G} , $Y \in \mathcal{F}$, then

$$E(XY|\mathcal{G}) = XE(Y|\mathcal{G}) \text{ a.s. } P$$

Other Properties

- (a) X, Y are random elements such that XY integrable P .
- (b) If $\mathcal{G} \subseteq \mathcal{F}$ is the sub σ -field, then

$$E(XE(Y|\mathcal{G})) = E(E(X|\mathcal{G})Y) = E(E(X|\mathcal{G})E(Y|\mathcal{G}))$$

Conditional expectation is a self-adjoint operation.

Proof. "Wire Theorem"

$$\begin{aligned} E(XE(Y|\mathcal{G})) &= E(E(XE(Y|\mathcal{G})|\mathcal{G})) \\ &= E(E(Y|\mathcal{G})E(X|\mathcal{G})) \\ &= E(E(E(X|\mathcal{G})Y|\mathcal{G})) \\ &= E(E(X|\mathcal{G})Y) \end{aligned}$$

■

8. Conditional Distribution of a Random Element Given Another Random Element

Here we have the typical probability space, measurable spaces for X and Y .

Let there be a function,

$$h : \mathcal{F}_X \times \Omega_Y \rightarrow \mathbb{R}$$

This function is called the conditional distribution of X given Y if

$$\tilde{h}(A, \omega) = h(A, Y(\omega))$$

We say that $\tilde{h} : \mathcal{F}_X \times \Omega_Y \rightarrow \mathbb{R}$ is the conditional distribution of X given $\mathcal{G} = Y^{-1}(\mathcal{F}_Y)$. That is,

- (a) For each $A \in \mathcal{F}_X$

$$\tilde{h}(A, Y(\cdot)) = P(X^{-1}(A) | Y^{-1}(\mathcal{F}_Y))$$

- (b) For each $\omega \in \Omega$

$$\tilde{h}(\cdot, Y(\omega)) = P_{X|Y}(A|y)$$

9. Conditional Density of One Random Element Given Another Random Element

Suppose probability space and σ -finite measure spaces for X and Y .

Here our relevant function is

$$g : \Omega_X \times \Omega_Y$$

which is the conditional density of X given Y if for all $A \in \mathcal{F}_X$,

$$\int_A g(x, y) d\mu_X(x) = P_{X|Y}(A|y)$$

In the following special case, g has an explicit formula.

$$(\Omega, \mathcal{F}, P)$$

$$(\Omega_X, \mathcal{F}_X, \mu_X)$$

$$(\Omega_Y, \mathcal{F}_Y, \mu_Y)$$

$$(\Omega_X \times \Omega_Y, \mathcal{F}_X \times \mathcal{F}_Y, \mu_X \times \mu_Y)$$

$$(X, Y) : \Omega \rightarrow \Omega_X \times \Omega_Y \subseteq \mathcal{F} \setminus \mathcal{F}_X \times \mathcal{F}_Y$$

$$\text{Let } P_X = PX^{-1}, P_Y = PY^{-1}, P_{XY} = P(XY)^{-1}.$$

$$\text{Assume } P_X \ll \mu_X, P_Y \ll \mu_Y, P_{XY} \ll \mu_X \times \mu_Y.$$

$$f_X = \frac{dP_X}{d\mu_X}$$

$$f_Y = \frac{dP_Y}{d\mu_Y}$$

$$f_{XY} = \frac{dP_{XY}}{d(\mu_X \times \mu_Y)}$$

Let

$$f_{X|Y} = \begin{cases} \frac{f_{XY}}{f_Y} & \text{if } f_Y \neq 0 \\ 0 & \text{if } f_Y = 0 \end{cases}$$

$$f_{Y|X} = \begin{cases} \frac{f_{XY}}{f_X} & \text{if } f_X \neq 0 \\ 0 & \text{if } f_X = 0 \end{cases}$$

Then it is easy to show that each is indeed the conditional density of their respective elements (first given second).

Wednesday January 18

Claim: $g(x, y)$ is the conditional density.

Proof. Want to show that for all $A \in \mathcal{F}_X$,

$$\int_A g(x, y) d\mu_X(x) = P_{X|Y}(A|y)$$

Which means that

$$\int_A g(x, y(\omega)) d\mu_X(x) = P_{X|Y}(X^{-1}(A) | \sigma(y))$$

This is true if for all $G' \in \sigma(y)$

$$\int_{G'} \int_A g(x, y(\omega)) d\mu_X(x) dP(\omega) = P(X^{-1}(A) \cap G')$$

But note that

$$\begin{aligned} G' &\in \sigma(y) \\ \Leftrightarrow G' &\in Y^{-1}(\mathcal{F}_Y) \\ G' &= Y^{-1}(G) \text{ for some } G \in \mathcal{F}_Y \end{aligned}$$

So we want to check that

$$\begin{aligned}
\int_{Y^{-1}(G)} \int_A g(x, y(\omega)) d\mu_X(x) dP(\omega) &= P(X^{-1}(A) \cap Y^{-1}(G)) \\
\int_{Y^{-1}(G)} \int_A g(x, y(\omega)) d\mu_X(x) dP(\omega) &= \int_G \int_A g(x, y) d\mu_X(x) dP_Y(y) \\
&= \int_G \int_A \frac{f_{XY}(x, y)}{f_Y(y)} d\mu_X(x) [f_Y(y)] d\mu_Y(y) \\
&= \int_G \int_A f_{XY}(x, y) d\mu_X(x) d\mu_Y(y) \\
&= \int_{G \times A} f_{XY}(x, y) d(\mu_X \times \mu_Y)(x, y) \\
&= P_{XY}(G \times A) \\
&= P \circ (X, Y)^{-1}(A \times G) \\
&= P(X \in A, Y \in G) \\
&= P(\omega : \omega \in X^{-1}(A) \& \omega \in Y^{-1}(G)) \\
&= P(X^{-1}(A) \cap Y^{-1}(G))
\end{aligned}$$

■

1.1 Frequentist & Bayesian Settings

We have our probability space (Ω, \mathcal{F}, P) . We also have some data,

$$\begin{aligned}
&(\Omega_X, \mathcal{F}_X, \mu_X) \\
X : \Omega &\rightarrow \Omega_X \text{ (mod) } \mathcal{F} / \mathcal{F}_X
\end{aligned}$$

Here, usually Ω_X is a \mathbb{R}^m .

Typically we have

$$X = (X_1, \dots, X_n)$$

and possibly,

$$X_i = \begin{pmatrix} X_{i1} \\ \vdots \\ X_{ip} \end{pmatrix}$$

We could say that these data are independent and identically distributed (iid) random vectors of dimension p . In this case $m = np$.

The goal of statistical inference is to estimate.

$$P_X = PX^{-1} = P_0$$

The ?? distribution of X .

There are two schools of thought

1. Frequentist Approach - assume a family of distributions, \mathcal{P} , where $\mathcal{P} \ll \mu_X$. Usually we assume that \mathcal{P} is a parametric family, $\mathcal{P} = \{P_\theta : \theta \in \Omega_\theta \subseteq \mathbb{R}^p\}$. We assume that $P_0 \in \mathcal{P}$, that is there exists $\theta_0 \in \Omega_\theta$ such that $P_\theta = P_0$. The goal is to estimate P_0 .
2. Bayesian Approach - here we assume the data is generated by the conditional distribution $P_{X|\theta}$. We observe X , then determine what is the best estimate of the random θ .

1.2 Prior Posterior & Likelihood

Here let there be probability space (Ω, \mathcal{F}, P) ; σ -finite measurable spaces $(\Omega_X, \mathcal{F}_X, \mu_X)$, $(\Omega_\theta, \mathcal{F}_\theta, \mu_\theta)$. Together,

$$(\Omega_X \times \Omega_\theta, \mathcal{F}_X \times \mathcal{F}_\theta, \mu_X \times \mu_\theta)$$

Also, a random element,

$$(X, \theta) : \Omega \rightarrow \Omega_X \times \Omega_\theta \subseteq \mathcal{F} \setminus \mathcal{F}_X \times \mathcal{F}_\theta$$

$P_X = P \circ X^{-1} \leftarrow$ marginal distribution of X
 $P_\theta = P \circ \theta^{-1} \leftarrow$ prior distribution
 $P_{X,\theta} = P \circ (X, \theta)^{-1} \leftarrow$ joint distribution of X and θ
 $P_{X|\theta}(A|\theta) : \mathcal{F}_X \times \Omega_\theta \rightarrow \mathbb{R}$. Likelihood distribution
 $P_{\theta|X}(G|x) : \mathcal{F}_\theta \times \Omega_X \rightarrow \mathbb{R}$. Posterior distribution

Note in the following the first inequalities are **assumed**.

$P_X \ll \mu_X \Rightarrow f_X = \frac{dP_X}{d\mu_X}$ Marginal Density
 $P_\theta \ll \mu_\theta \Rightarrow \pi_\theta = \frac{dP_\theta}{d\mu_\theta}$ Prior Density
 $P_{X,\theta} \ll \mu_X \times \mu_\theta \Rightarrow f_{X,\theta}(x, \theta) = \frac{dP_{X,\theta}}{d(\mu_X \times \mu_\theta)}$ Joint Density

FINISH FROM PHOTO

One way to estimate θ is by maximizing $\pi_{\theta|X}(\theta|x)$. Want to do so with value that is most likely to happen (given the data).

$$\pi_{\theta|X}(\theta|x) = P_{\theta|X}(\theta = \theta|x)$$

By construction,

$$\begin{aligned} \pi_{\theta|X} &= \frac{f_{X,\theta}}{f_X} \\ &= \frac{f_{X|\theta} \pi_\theta}{\int_{\Omega_\theta} f_{X|\theta} \pi_\theta d\mu_\theta} \end{aligned}$$

1.3 Conditional Independence and Frequentist/Bayesian Sufficiency

Independence

Two random elements are said to be independent if for all $A' \in \sigma(X)$, $G' \in \sigma(\theta)$ we have

$$P(A' \cap G') = P(A')P(G')$$

This statement can also be expressed in $(\Omega_X \times \Omega_\Theta, \mathcal{F}_X \times \mathcal{F}_\Theta, P_X \times P_\Theta)$ as follows.

Since $A' \in \sigma(X) = X^{-1}(\mathcal{F}_X)$, $A' = X^{-1}(A)$ for some $A \in \mathcal{F}_X$. So $G' = \Theta^{-1}(G)$, $G \in \mathcal{F}_\Theta$.

$$\begin{aligned}
 P(A' \cap G') &= P(X^{-1}(A) \cap \Theta^{-1}(G)) \\
 &= P(\{\omega : \omega \in X^{-1}(A) \cap \Theta^{-1}(G)\}) \\
 &= P(\{\omega : \omega \in X^{-1}(A) \& \omega \in \Theta^{-1}(G)\}) \\
 &= P(\{\omega : X(\omega) \in A, \Theta(\omega) \in G\}) \\
 &= P(\{\omega : (X(\omega), \Theta(\omega)) \in AxG\}) \\
 &= P(\{\omega : (X, \Theta)(\omega) \in AxG\}) \\
 &= P(\{\omega : \omega \in (X, \Theta)^{-1}AxG\}) \\
 &= [P \circ (X, \Theta)^{-1}](AxG) \\
 &= P_{X, \Theta}(AxG)
 \end{aligned}$$

Also note that

$$P(A') = P(X^{-1}(A)) = P_X(A)$$

$$P(G') = P_\Theta(G)$$

So with independence, (and for $A \in \mathcal{F}_X, G \in \mathcal{F}_\Theta$)

$$P_{X, \Theta}(AxG) = P_X(A)P_\Theta(G)$$

But we know that this implies that $P_{X, \Theta}$ is the product measure $P_X \times P_\Theta$.

Conditional Independence

Now, given sub σ -field $\mathcal{G} \in \mathcal{F}$ we want to define $X \& \Theta$ conditionally independent given \mathcal{G} .

Definition 1.3.1 We say that $X \& \Theta$ are conditionally independent given \mathcal{G} (i.e. $X \perp\!\!\!\perp \Theta | \mathcal{G}$) if for all $A' \in \sigma(X), G' \in \sigma(\Theta)$ we have

$$P[A' \cap G' | \mathcal{G}] = P[A' | \mathcal{G}]P[G' | \mathcal{G}] \text{ a.s. } P$$

Equivalently for all $A \in \mathcal{F}_X, G \in \mathcal{F}_\Theta$,

$$P[X^{-1}(A) \cap \Theta^{-1}(G) | \mathcal{G}] = P[X^{-1}(A) | \mathcal{G}]P[\Theta^{-1}(G) | \mathcal{G}]$$

Equivalently,

$$P_{X, \Theta | \mathcal{G}}(AxG | \mathcal{G}) = P_{X | \mathcal{G}}(A | \mathcal{G})P_{\Theta | \mathcal{G}}(G | \mathcal{G})$$

Equivalent Condition for Conditional Independence

Theorem 1.3.1 — 1.1 in Notes. The following statements are equivalent.

1. $X \perp\!\!\!\perp \Theta | \mathcal{G}$
2. $P(X^{-1}(A) | \Theta, \mathcal{G}) = P(X^{-1}(A) | \mathcal{G}) \text{ a.s. } P \quad \forall A \in \sigma(X)$
3. $P(\Theta^{-1}(G) | X, \mathcal{G}) = P(\Theta^{-1}(G) | \mathcal{G}) \text{ a.s. } P \quad \forall G \in \sigma(\Theta)$

Proof. It suffices to proof that $1 \Leftrightarrow 2$.

$1 \Rightarrow 2$. We know that for all $A \in \mathcal{F}_X, G \in \mathcal{F}_\Theta$ that

$$P[X^{-1}(A) \cap \Theta^{-1}(G) | \mathcal{G}] = P[X^{-1}(A) | \mathcal{G}] P[\Theta^{-1}(G) | \mathcal{G}]$$

Want that for all $A \in \mathcal{F}_X$ that $P(X^{-1}(A) | \Theta, \mathcal{G}) = P(X^{-1}(A) | \mathcal{G})$.

$$\begin{aligned} P(X^{-1}(A) | \Theta, \mathcal{G}) &\equiv P(X^{-1}(A) | \sigma(\sigma(\Theta) \cup \mathcal{G})) \\ &= P(\dots | \sigma(\Theta^{-1}(\mathcal{F}_\Theta) \cup \mathcal{G})) \end{aligned}$$

So it suffices to show that

$$P(X^{-1}(A) | \sigma(\Theta^{-1}(\mathcal{F}_\Theta) \cup \mathcal{G})) = P(X^{-1}(A) | \mathcal{G})$$

From the definition given we want to show that the above statement is true. which is so that the for all $B \in \sigma(\Theta^{-1}(\mathcal{F}_\Theta) \cup \mathcal{G})$,

$$\int_B P(X^{-1}(A) | \mathcal{G}) dP = P(X^{-1}(A) \cap B)$$

But this is very hard because B is hard to characterize. But we have theorem that says you only have to check (*) for all B in a π -system generating $\sigma(\Theta^{-1}(\mathcal{F}_\Theta) \cup \mathcal{G})$.

$$\mathcal{P} = \{\Theta^{-1}(G) \cap F : G \in \mathcal{F}_\Theta, F \in \mathcal{G}\}$$

It is trivial to show that \mathcal{P} is a π -system.

MORE IN PHOTO

Meanwhile,

$$\mathcal{P} \subseteq \sigma(\Theta^{-1}(\mathcal{F}_\Theta) \cup \mathcal{G})$$

Therefore,

$$\sigma(\Theta^{-1}(\mathcal{F}_\Theta) \cup \mathcal{G}) = \sigma(\mathcal{P})$$

So, sufficient to check (*) $\forall B \in \mathcal{P}'$

$$B \in \mathcal{P} \Rightarrow B = \Theta^{-1}(G) \cap F, G \in \mathcal{F}_\Theta, F \in \mathcal{G}$$

So, we want

$$\int_{\Theta^{-1}(G) \cap F} P(X^{-1}(A) | \mathcal{G}) dP = P(\Theta^{-1}(G) \cap F \cap X^{-1}(A))$$

$$\begin{aligned}
\int_{\Theta^{-1}(G) \cap F} P(X^{-1}(A) | \mathcal{G}) dP &= \int_{\Theta^{-1}(G) \cap F} E(I_{X^{-1}(A)} | \mathcal{G}) dP \\
&= E(I_{\Theta^{-1}(G)} I_F E(I_{X^{-1}(A)} | \mathcal{G})) \\
&= E(E(I_{\Theta^{-1}(G)} I_F | \mathcal{G}) E(I_{X^{-1}(A)} | \mathcal{G})) \\
&= E(I_F E(I_{\Theta^{-1}(G)} | \mathcal{G}) E(I_{X^{-1}(A)} | \mathcal{G})) \\
&= E(I_F E(I_{\Theta^{-1}(G)} I_{X^{-1}(A)} | \mathcal{G})) \\
&= E(E(I_F I_{\Theta^{-1}(G)} I_{X^{-1}(A)} | \mathcal{G})) \\
&= E(I_F I_{\Theta^{-1}(G)} I_{X^{-1}(A)}) \\
&= P(F \cap \Theta^{-1}(G) \cap X^{-1}(A))
\end{aligned}$$

Monday January 23

$2 \Rightarrow 1$. We want to show that

$$P(X^{-1}(A) | \mathcal{G}) P(\Theta^{-1}(G) | \mathcal{G})$$

is conditional probability of

$$P(X^{-1}(A) \cap \Theta^{-1}(G) | \mathcal{G})$$

for all $F \in \mathcal{G}$.

$$\begin{aligned}
\int_F P(X^{-1}(A) | \mathcal{G}) P(\Theta^{-1}(G) | \mathcal{G}) dP &= E[I_F E(I_{X^{-1}(A)} | \mathcal{G}) E(I_{\Theta^{-1}(G)} | \mathcal{G})] \\
&= E[E(I_{X^{-1}(A)} | \mathcal{G}) E(I_F I_{\Theta^{-1}(G)} | \mathcal{G})] \\
&= E[E(I_{X^{-1}(A)} | \mathcal{G}) I_F I_{\Theta^{-1}(G)}] \\
&= E[E(I_{X^{-1}(A)} I_F I_{\Theta^{-1}(G)} | \Theta, \mathcal{G})] \\
&= E[I_{X^{-1}(A)} I_F I_{\Theta^{-1}(G)}] \\
&= P(X^{-1}(A) \cap \Theta^{-1}(G) \cap F)
\end{aligned}$$

■

1.4 Equivalence of Frequentist & Bayesian Sufficiency

Here we have,

$$(\Omega_\Theta, \mathcal{F}_\Theta, \mu_\Theta), (\Omega_X, \mathcal{F}_X, \mu_X), (\Omega_T, \mathcal{F}_T)$$

Where

$$T : \Omega_X \rightarrow \Omega_T \text{ } \mathbb{M} \mathcal{F}_X / \mathcal{F}_T$$

is called a statistic.

$$T = T(X) \text{ or } T \circ X = T(X(\omega))$$

In frequentist setting, we say that T is **sufficient** if $P_{X|T,\Theta}$ does not depend on Θ . It can be easily verified (see Homework) that $P_{X|T,\Theta}$ doesn't depend on Θ implies that

$$P_{X|T,\Theta} = P_{X|T} \text{ a.s. } P$$

This is "nearly" frequentist. Above is exchangeable with " $X \perp\!\!\!\perp \Theta | T$ ", but can't say this in frequentist setting.

$$P_{\Theta|T,X} = P_{\Theta|T} \Leftrightarrow P_{\Theta|X} = P_{\Theta|T}$$

That is to say that a statistic, T , is sufficient for Θ if and only iff the posterior distribution of $\Theta|X$ is the same as the posterior distribution of $\Theta|T$. This would be used in a Bayesian setting.

Definition 1.4.1 — Bayesian Sufficient. We say that $T \circ X$ is **Bayesian sufficient** if

$$P_{\Theta|X} = P_{\Theta|T} \text{ a.s. } P$$

Lemma 1.1 (HW 2) Suppose that $f(\theta)$ is a p.d.f such that

$$f(\theta) \propto \exp\{-a\theta^2 + b\theta\}, \quad a > 0$$

Then,

1. $\theta \sim N(\frac{b}{2a}, \frac{1}{2a})$
2. $\int \exp\{-a\theta^2 + b\theta\} d\theta = \sqrt{\frac{\pi}{a}} \exp\{\frac{b^2}{4a}\}$

■ **Example 1.1** Suppose that

$$\begin{aligned} X|\Theta &\sim N(\Theta, \sigma^2) \\ \Theta &\sim N(\mu, \tau^2) \end{aligned}$$

Find $\pi_{\Theta|X}(\theta|x), f_X(x)$.

Solution:

$$\begin{aligned} \pi(\theta|x) &\propto f(x|\theta)\pi(\theta) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \frac{(x-\theta)^2}{\sigma^2}\right\} * \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2} \frac{(\theta-\mu)^2}{\tau^2}\right\} \\ &\propto \exp\left\{-\frac{1}{2} \frac{(x-\theta)^2}{\sigma^2}\right\} * \exp\left\{-\frac{1}{2} \frac{(\theta-\mu)^2}{\tau^2}\right\} \\ &= \exp\left\{-\left(\frac{1}{2\sigma^2} + \frac{1}{2\tau^2}\right)\theta^2 + \left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)\theta\right\} \end{aligned}$$

Using Lemma 1.1,

$$\theta|X \sim N\left(\frac{\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}}{1/2(2\sigma^{-2} + 1/2\tau^{-1})}, \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\tau^2}}\right)$$

How about $f_X(x)$?

$$\begin{aligned}
f_X(x) &= \int f(x|\theta)\pi(\theta)d\theta \\
&\vdots \\
&= \frac{1}{2\pi\sigma\tau} * \exp\left\{-\frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2}\right\} \int \exp\left\{-\left(\frac{1}{2\sigma^2} + \frac{1}{2\tau^2}\right)\theta^2 + \left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)\theta\right\} \\
&= \dots * \sqrt{\frac{\pi}{\frac{1}{2\sigma^2} + \frac{1}{2\tau^2}}} \exp\left\{\frac{\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)^2}{4\left(\frac{1}{2\sigma^2} + \frac{1}{2\tau^2}\right)}\right\}
\end{aligned}$$

We want to identify this as a p.d.f of x , so we can treat anything that is not x as a constant. Using elementary algebra we get...

$$\propto \exp\left\{-\left(\frac{x^2}{2(\tau^2 + \sigma^2)} + \frac{x\mu}{(\sigma^2 + \tau^2)}\right)\right\}$$

Applying Lemma 1.1 for x and simplifying,

$$X \sim N(\mu, \tau^2 + \sigma^2)$$

■

This can be extended to multivariate setting, 2-sample setting, ANOVA setting, regression setting, etc. It is essential to all aspects of linear models.

Wednesday January 25

■ **Example 1.2** Suppose

$$\begin{aligned}
X_1, \dots, X_n | \theta &\stackrel{iid}{\sim} N(\theta, \sigma^2) \\
\theta &\sim N(\mu, \tau^2)
\end{aligned}$$

$$\pi(\theta | X_1, \dots, X_n) = ?$$

By Example 1.1,

$$\theta | \bar{X} \sim N\left(\frac{\frac{\bar{X}}{\sigma^2/n} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2/n} + \frac{1}{\tau^2}}, \frac{1}{\frac{1}{\sigma^2/n} + \frac{1}{\tau^2}}\right)$$

Note that the sample size will effect how much weight each of the pervious means is given. As $n \rightarrow \infty$, depending on which coeffercient goes to 1,

$$E(\theta | \underline{X}) \rightarrow \bar{X} \text{ or } \mu$$

Similarly with variance, as $n \rightarrow \infty$,

$$\text{Var}(\theta | \underline{X}) \rightarrow 0$$

We can generalize/approximate this phenomenon as follows,

$$\theta | X_1, \dots, X_n \sim N(\hat{\theta}_1, I^{-1}(\hat{\theta}))$$

where $\hat{\theta}$ is the MLE.

In our special case,

$$\theta|X_1, \dots, X_n \sim N(\bar{X}, \frac{\sigma^2}{n})$$

■

1.5 Conjugate Priors

1.5.1 Introduction

In general, computing posterior distributions or posterior means is a hard problem involving high-dimensional numerical integration. This was a big hurdle for Bayesian methods before computers. Now we can do methods such as Monte Carlo integration (MCMC).

In special cases, such as with the Exponential Family and mixture of distributions, posterior can be expressed explicitly through the use of conjugate families.

Definition 1.5.1 — Conjugate Family. A family of distributions, \mathcal{P} , on $(\Omega_\theta, \mathcal{F}_\theta)$ is a **conjugate family** if

$$P_\theta \in \mathcal{P} \Rightarrow \mathcal{P}_{\theta|X} \in \mathcal{P}$$

Not unique, for example if you let \mathcal{P} be the collection of all distributions on $(\Omega_\theta, \mathcal{F}_\theta)$, then it is always conjugate. Usually there is a suitable conjugate family.

1.5.2 Exponential Family

■ **Example 1.3** $X_1, \dots, X_n | \theta \sim \text{Pois}(\theta)$

$$\begin{aligned} f(x_1, \dots, x_n | \theta) &= \prod_{i=1}^n \frac{\theta^{x_i}}{x_i!} e^{-\theta} \\ &= \frac{\theta^{\sum x_i} e^{-n\theta}}{\prod (x_i!)} \\ &\propto \theta^{c_1} e^{-c_2 \theta} \end{aligned}$$

Note that we are treating θ as the variable of interest here, not x . Recall that if $\theta \sim \text{Gamma}(\alpha, \beta)$ then we have a distribution approaching the form above, that is,

$$\pi(\theta) \propto \theta^{\alpha-1} e^{-\theta/\beta}$$

If we use this form, then we may find $pi(\theta|\underline{X})$ using the following:

$$pi(\theta|\underline{X}) \propto \theta^{c_1+\alpha-1} e^{-(c_2+1/\beta)\theta}$$

Which gives us that

$$\theta|X_1, \dots, X_n \sim \text{Gamma}(c_1 + \alpha, (c_2 + 1/\beta)^{-1})$$

This is generally true for all exponential family distributions.

■

We say that X has exponential family distribution if it has p.d.f in the form of

$$\frac{e^{\theta^T t(x)}}{\int e^{\theta^T t(x)} d\mu(x)}$$

where μ is a σ -finite measure on $(\Omega_X, \mathcal{F}_X)$.

Essentially, p.d.f. with $\mu(x) \propto e^{\theta^T t(x)}$.

More generally, suppose that $\phi : \Theta \mapsto \Phi$ bijection (one-to-one onto). Then X has Exponential Family distribution if and only if the p.d.f of X with μ is

$$\frac{e^{\phi^T t(x)}}{\int e^{\phi^T t(x)} d\mu(x)}$$

In this case, $X \sim \text{Exp}(\phi, t, \mu)$, when ϕ is identity, this is called the canonical form of exponential family.

Theorem 1.5.1 — 1.3 from class. If

$$P_\Theta \sim \text{Ep}(\xi, \phi, \nu)$$

$$P_{X|\Theta} \sim \text{Ex}(\phi, t, \mu)$$

then

$$P_{\Theta|X} \in \text{Ep}(\xi_X, \phi, \nu_X)$$

where

$$\xi_X(\alpha) = \xi(\alpha) + t(x)$$

$$d\nu_X(\theta) = \frac{d\nu(\theta)}{\int e^{\phi^T(\theta)t(x)} d\mu(x)}$$

$$\text{Proof. } P_\Theta \in \text{Ep}(\xi, \phi, \nu) \Rightarrow \pi(\theta) = \frac{e^{\xi^T(\alpha)t(\theta)}}{\int e^{\xi^T(\alpha)t(\theta)} d\nu(x)}$$

$$f(x|\theta) = \frac{e^{\phi^T t(x)}}{\int e^{\phi^T t(x)} d\mu(x)}$$

$$\pi(\theta|x) = \text{PHOTO}$$

$$=$$

GIANT fraction of fractions = PHOTO

So, with respect to the new measure,

$$d\nu_X(\theta) = \frac{d\nu(\theta)}{\int e^{\phi^T(\theta)t(x)} d\mu(x)}$$

where the pdf is

So, $\theta|X \sim Ep(\xi(\alpha) + t(x), \phi(\theta), \frac{dv(\theta)}{\int e^{\phi^T(\theta)t(x)d\mu(x)})}$

■

Friday January 27

■ **Example 1.4 — One Parameter Normal.**

$$X|\theta \sim N(\theta, \sigma^2)$$

Want to assign conjugate prior for θ .

$$f(x|\theta) \propto \exp\left\{\left(\frac{-1}{2\sigma^2}\right)\theta^2 + \left(\frac{\mu}{\sigma^2}\right)\theta\right\}$$

Recall Lemma 1.1.

$$f(\theta) \propto \exp\left\{\left(\frac{-1}{2\sigma^2}\right)\theta^2 + \left(\frac{\mu}{\sigma^2}\right)\theta\right\}$$

$$\Rightarrow \theta \sim N(\mu, \sigma^2)$$

So, consider the following family,

$$\mathcal{F} = \left\{f(\theta) \propto \exp\left\{\frac{-1}{2\alpha_2^2}\theta^2 + \frac{\alpha_1}{\alpha_2^2}\theta\right\}\right\}$$

Suppose that $\pi \in \mathcal{F}$. Then we have that

$$\begin{aligned} \pi(\theta|X) &\propto f(x|\theta)\pi(\theta) \\ &\propto \exp\left\{\left(\frac{-1}{2\sigma^2}\right)\theta^2 + \left(\frac{\mu}{\sigma^2}\right)\theta\right\} * \exp\left\{\frac{-1}{2\alpha_2^2}\theta^2 + \frac{\alpha_1}{\alpha_2^2}\theta\right\} \\ &= \exp\left\{\left(\frac{-1}{2\alpha_2^2} - \frac{1}{2\alpha_2^2}\right)\theta^2 + \left(\frac{x}{\sigma^2} + \frac{\alpha_1}{\alpha_2^2}\right)\theta\right\} \\ &= -\frac{1}{2}\left(\frac{1}{\sigma^2} + \frac{1}{\alpha^2}\right) \end{aligned}$$

So we have that

$$\theta|X \sim N\left(\frac{\frac{\alpha_1}{\alpha_2} + \frac{x}{\sigma^2}}{\left(\frac{1}{\sigma^2} + \frac{1}{\alpha^2}\right)}, \frac{1}{\sigma^2 + \frac{1}{\alpha^2}}\right)$$

■

1.5.3 Convex Hull of Conjugate Family



Mathematically, if $A \subseteq \mathbb{R}^k$,

$$\text{Conv}(A) = \{\alpha_1 s_1 + \dots + \alpha_k s_k : \alpha_1 \geq 0, \dots, \alpha_k \geq 0; \alpha_1 + \dots + \alpha_k = 1; s_1 \in A, \dots, s_k \in A\}$$

R By the way, a convex combination of a set of probability measures, say P_1, \dots, P_k , is called the **mixture** of P_1, \dots, P_k . So $\text{Conv}(\mathcal{P})$ is simply the collection of all mixture distributions derived from \mathcal{P} .

Theorem 1.5.2 If \mathcal{P} is conjugate to $P_{X|\theta}$, then $\text{Conv}(\mathcal{P})$ is also conjugate.

Proof. Suppose that $P_\Theta \in \text{Conv}(\mathcal{P})$. Want to show that $P_{\Theta|X} \in \text{Conv}(\mathcal{P})$.

Since $P_\Theta \in \text{Conv}(\mathcal{P})$ there exists

$$\alpha_1, \dots, \alpha_k \in \mathbb{R}, \sum_{i=1}^k \alpha_i = 1, \alpha_i \geq 0$$

$$P_\Theta^{(1)} \dots P_\Theta^{(k)} \in \mathcal{P}$$

such that

$$P_\Theta = \sum_{i=1}^k \alpha_i P_\Theta^{(i)}$$

$$\begin{aligned} f_X(x) &= \int f(x|\theta) dP_\Theta \\ &= \int f(x|\theta) d\left(\sum_{i=1}^k \alpha_i P_\Theta^{(i)}\right) \\ &= \sum_{i=1}^k \alpha_i \int f(x|\theta) dP_\Theta^{(i)} \\ &= \sum_{i=1}^k \alpha_i \int m_i(x) dP_\Theta^{(i)} \\ dP_\Theta(\cdot|x) &= \frac{f(x|\theta)}{f_X(x)} dP_\Theta \\ &= \frac{f(x|\theta)}{\sum_{i=1}^k \alpha_i m_i(x)} \sum_{i=1}^k \alpha_i dP_\Theta^{(i)} \\ &= \sum_{i=1}^k \frac{\alpha_i m_i(x)}{\sum_{j=1}^k \alpha_j m_j(x)} \frac{f(x|\theta) dP_\Theta^{(i)}}{m_i(x)} \end{aligned}$$

Note that the first term is greater than zero and sums to 1. It's our new α^* . In the second term note that $P_\Theta^{(i)} \in \mathcal{P}$ and $P_{\Theta|X}^{(i)}(\cdot|x) \in \mathcal{P}$.

Thus this is a convex combination of members of \mathcal{P} in $\text{Conv}(\mathcal{P})$. ■

This is a nice way to construct conjugate families of mixtures of exponential family.

1.6 Two-Parameter, Normal Family

Definition 1.6.1 — Inverse χ^2 Distribution. A random variable T is said to have an inverse χ^2 distribution of $\chi_{(k)}^{-2}$ if and only if,

$$\frac{1}{T} \sim \chi_{(k)}^2$$

If, for some $\tau < 0$, $\frac{T}{\tau} \sim \chi_{(k)}^{-2}$ then we write that

$$T \sim \tau \chi_{(k)}^{-2}$$

By Jacobian theorem,

$$T = g(X)$$

$$f_T(t) = f_X(g^{-1}(t)) \left| \det \left(\frac{\partial g^{-1}(t)}{\partial t} \right) \right|$$

Above will be shown in HW.

Theorem 1.6.1 If $T \sim \chi_{(v)}^{-2}$ then,

$$f(t) = \frac{1}{\Gamma(\frac{v}{2}) 2^{\frac{v}{2}}} t^{-\frac{v}{2}-1} e^{-\frac{t}{2}}, t > 0$$

Proof. Shown in HW 2. ■

Now suppose that $X_1, \dots, X_n | \phi, \lambda \stackrel{iid}{\sim} N(\lambda, \phi)$ where both λ and ϕ are random. How do we assign conjugate prior of (λ, ϕ) ?

From classical statistics (Stat 514) we know that the sufficiency statistic for λ, ϕ is $(\sum X_i, \sum X_i^2)$ or $(T = \bar{X}, S = \sum_{i=1}^n (X_i - \bar{X})^2)$.

So $\pi(\lambda, \phi | X_1, \dots, X_n) = \pi(\lambda, \phi | T, S)$. Moreover, from Normal theory,

$$T \perp\!\!\!\perp X | \lambda, \phi$$

$$T | \lambda, \phi \sim N(\lambda, \phi/n)$$

$$S | \lambda, \phi \sim \phi \chi_{(n-1)}^2$$

Monday January 30

So, we need only to look at [GET NOTES FROM BUDDY]

$$\begin{aligned} f(T, S | \lambda, \phi) &= f(S | \lambda, \phi) f(T | \lambda, \phi) \\ &= f(S | \phi) f(T | \lambda, \phi) \\ &= [S | \phi] [T | \lambda, \phi] \\ &= \left[\phi^{-(\frac{n-1}{2}-1)} \exp \left\{ \left(\frac{-S}{2\phi} \right) \left(\frac{1}{\phi} \right) \right\} \right] * \left[\frac{1}{\sqrt{2\pi}(\phi/n)^{1/2}} \exp \left\{ \frac{-1}{2(\phi/n)} \lambda^2 + \frac{t}{\phi/n} t + \left(\frac{-1}{2(\phi/n)} \right) t^2 \right\} \right] \end{aligned}$$

So we have $N(t, \frac{\phi}{n}) s \chi_{(n-3)}^{-2}$.

More generally, the distribution,

$$N(a, \frac{\phi}{m}) * \tau \chi_{(k)}^{-2}$$

is referred to (by Bing Li) as NICH (Normal-Inverse CHi-square). Here,

$$NICH(a, m, \tau, k)$$

We may say that $(\lambda, \phi) \mapsto [S | \phi] [T | \lambda, \phi] \sim NICH(t, n, s, n-3)$. The likelihood is NICH.

Lemma 1.2

$$NICH(a_1, m_1, \tau_1, k_1) * NICH(a_2, m_2, \tau_2, k_2) = NICH(a_3, m_3, \tau_3, k_3)$$

Where we have that

$$\begin{aligned} a_3 &= \frac{m_1 a_1 + m_2 a_2}{m_1 + m_2} \\ m_3 &= m_1 + m_2 \\ \tau_3 &= \tau_1 + \tau_2 + m_1 a_1^2 + m_2 a_2^2 - m_3 a_3^2 \\ k_3 &= k_1 + k_2 - 3 \end{aligned}$$

Before proof, let's rewrite likelihood as

$$NICH(a, m, \tau, k) \propto \phi^{-1/2} \exp\left\{\frac{-1}{2(\phi/m)} \lambda^2 + \frac{a}{\phi/m} \lambda\right\} \phi^{-(k/2)-1} \exp\left\{-\frac{\tau + m a^2}{2\phi}\right\}$$

Proof. By definition

$$\begin{aligned} NICH_1 * NICH_2 &\propto \phi^{-1/2} \exp\left\{\frac{-1}{2(\phi/m_1)} \lambda^2 + \frac{a_1}{\phi/m_1} \lambda\right\} \phi^{-(k_1/2)-1} \exp\left\{-\frac{\tau_1 + m_1 a_1^2}{2\phi}\right\} * \phi^{-1/2} \exp\left\{\frac{-1}{2(\phi/m_2)} \lambda^2 + \frac{a_2}{\phi/m_2} \lambda\right\} \phi^{-(k_2/2)-1} \exp\left\{-\frac{\tau_2 + m_2 a_2^2}{2\phi}\right\} \\ &= \phi^{-1/2} \exp\left\{\left(\frac{-1}{2(\phi/m_1)} - \frac{1}{2(\phi/m_2)}\right) \lambda^2 + \left(\frac{a_1}{\phi/m_1} + \frac{a_2}{\phi/m_2}\right) \lambda\right\} \phi^{-(k_1/2)-1} \phi^{-(k_2/2)-1} \exp\left\{-\frac{\tau_1 + \tau_2 + m_1 a_1^2 + m_2 a_2^2 - m_3 a_3^2}{2\phi}\right\} \\ &= \phi^{-1/2} \exp\left\{\left(\frac{-1}{2(\phi/m_3)}\right) \lambda^2 + \left(\frac{a_3}{\phi/m_3}\right) \lambda\right\} \phi^{-(k_3/2)-1} \exp\left\{-\frac{\tau_3 + m_3 a_3^2}{2\phi}\right\} \end{aligned}$$

■

Recall we have that $(\lambda, \phi) \mapsto [S|\phi][T|\lambda, \phi] \sim NICH(t, n, s, n-3)$. By Lemma 1.2, we assign prior

$$[\lambda|\phi][\phi] \sim NICH(a, m, \tau, k)$$

This gives us that

$$\pi(\lambda, \phi|T, S) \sim NICH(a^*, m^*, \tau^*, k^*)$$

where we have that

$$\begin{aligned} a^* &= \frac{ma + nT}{m+n} \\ m^* &= m + n \\ \tau^* &= \tau + S + ma^2 + nt^2 - (m+n)\left(\frac{ma+nt}{m+n}\right)^2 \\ k^* &= k + n - 3 + 3 \end{aligned}$$

So, $[\lambda|\phi, X] \sim N(a^*, \frac{\phi}{m^*})$, and $[\phi|X] \sim \tau^* \chi_{k^*}^{-2}$.

Note that we can make inference about λ, ϕ using

$$\frac{\lambda - a^*}{\sqrt{\phi/m^*}} | \phi, X \sim N(0, 1)$$

since the RHS doesn't depend on ϕ we have that it is independence form $\phi|X$.

$$\frac{\lambda - a^*}{\sqrt{\phi/m^*}} | X \sim N(0, 1)$$

Wednesday February 1

We want to make inference about

$$\frac{\frac{\lambda - a^*}{\sqrt{\phi/m^*}}}{\sqrt{\frac{\tau^*}{\phi}/m^*}} | X \sim t_{m^*}$$

We may reorganize this to be

$$\frac{m^*(\lambda - a^*)}{\sqrt{\tau^*/m^*}} \sim t_{m^*}$$

Similarly, for making inference about ϕ ,

$$\phi | X \sim \tau^* \chi_{m^*}^{-2}$$

$$\frac{\tau^*}{\phi} | X \sim \chi_{m^*}^2$$

1.7 Noninformative Prior

1.7.1 General Concept

Even when you don't have a prior distribution it's still beneficial to use Bayesian setting. For example, in dealing with nuisance parameters we deal with high dimension prior.

So here, try to use Bayesian to solve "no prior information" problem. We want to use a flat prior, e.g. the Lebesgue measure. But, in what is this flat? Suppose we impose the Lebesgue measure on θ . Then the prior for monotone transformation of θ , say θ^3 , is not Lebesgue anymore.

Definition 1.7.1 — Improper Prior. An infinite, but σ -finite measure on Ω_Θ is called an **improper prior**.

■ Example 1.5

$$X | \theta \sim N(\theta, \phi)$$

Note that here ϕ is known.

$$\pi(\theta) \equiv 1$$

$$f(X | \theta) = \sqrt{2\pi}^{-1} e^{-1/2*(x-\theta)^2}$$

$$\pi(\theta | X) = \sqrt{2\pi}^{-1} e^{-1/2*(x-\theta)^2} * 1 = N(X, 1)$$

$$f_X(x) = \frac{f(x|\theta)\pi(\theta)}{\pi(\theta|X)} = \frac{f(x|\theta)}{\pi(\theta|x)} = 1$$

The marginal is Lebesgue improper! But, posterior is important

■

■ Example 1.6

$$X_1, \dots, X_n | \lambda, \phi \sim N(\lambda, \phi)$$

Here both parameters are random.

We have sufficient statistics, $(S = \bar{X}, T = \sum (X_i - \bar{X})^2)$.

$$\begin{aligned} [(\lambda, \phi) \mapsto f(t, s | \lambda, \phi)] &\sim NICH(t, n, s, n-3) \\ &\propto \phi^{-1/2} \exp \left\{ \frac{-1}{2(\phi/n)} \lambda^2 + \frac{t}{\phi/n} \lambda \right\} \phi^{-((n-3)/2)-1} \exp \left\{ -\frac{s + nt^2}{2\phi} \right\} \end{aligned}$$

$$[\lambda | \phi] = 1$$

$$[\phi] = \frac{1}{\phi}$$

When we multiply the p.d.f of our NICH by $\frac{1}{\phi}$ we get NICH(t, n, s, n-1) following the same argument in Section 1.5 (?) (proper case).

We can show that

$$\begin{aligned} \frac{\sqrt{n}(\lambda - t(x))}{\sqrt{s(x)/(n-1)}} | X &\sim t_{(n-1)} \\ \frac{s(x)}{\phi} | X &\sim \chi^2_{(n-1)} \end{aligned}$$

Exactly the same as frequentist sample distribution, except what random has changed. ■

1.7.2 Invariant Prior

What is flat? What is a natural generalization of Lebesgue Measure?

Lebesgue measure is invariant under translation.

$$\theta \mapsto \theta + c$$

If λ is Lebesgue and T is translation,

$$\lambda \cdot T^{-1} = \lambda$$

This is natural generalization of flatness. Change T to be some other transformation that somehow resembles translation = group of transformation.

R Review definition of group of transformations.
 Ω (set)

\mathcal{G} is a set of bijections (one-to-one on to) on Ω such that

1. for all $g_1, g_2 \in \mathcal{G}$, $g_1 \circ g_2 \in \mathcal{G}$
2. for all $e \in \mathcal{G}$ such that $e \circ g = g \circ e = g$ for all $g \in \mathcal{G}$

$$3. g \in \mathcal{G} \Rightarrow g^{-1} \in \mathcal{G}$$

With our own problem,

$$\Omega_{\Theta} \subseteq \mathbb{R}^P$$

Consider a parametric group,

$$\{g_t : t \in \Omega_{\Theta}\}$$

Consider two types of transformations,

$$\theta \mapsto g_t(\theta) = L_t$$

$$\theta \mapsto g_{\theta}(t) = R_t$$

We may refer to these transformations as Left and Right transformations, respectively.

Two types of invariant priors or two generalizations of Lebesgue measure, 2-generalization of flatness.

Definition 1.7.2 A measure, Π , on Ω_{Θ} is the left Haar measure.

$$\Pi = \Pi \circ L_t^{-1} \forall t \in \Omega_{\Theta}$$

A measure, Π , is the right Harr measure if

$$\Pi = \Pi \circ R_t^{-1} \forall t \in \Omega_{\Theta}$$

■ **Example 1.7**

$$\Omega_{\Theta} = \mathbb{R}$$

Translation group;

$$\mathcal{G} = \{(\theta \mapsto \theta + c = g_c(\theta)) : c \in \mathbb{R}\}$$

Suppose that the improper prior density of θ is $\pi(\theta)$.

$$L_t(\theta) = g_c(\theta) = \theta + c = \eta$$

then the density of η is

$$\phi(g_c^{-1}(\eta)) \left| \frac{\partial g_c^{-1}(\eta)}{\partial \eta} \right|$$

We want to find

$$\pi(\cdot) = \pi(\cdot - c)$$

$$\pi(\theta) = \pi(\theta - c) \quad \forall \theta \in \mathbb{R}$$

If we take $\theta = 0$,

$$\pi(0) = \pi(-c)$$

If we take $-c = \theta$,

$$\pi(\theta) = \pi(0) \propto \text{Lebesgue}$$

Left Haat measure IS the Lebesgue measure.

Right transformation,

$$\begin{aligned}
 R_c(\theta) &= g_\theta(c) \\
 &= c + \theta \\
 &= \theta + c \\
 &= g_c(\theta) \\
 &= L_c(\theta)
 \end{aligned}$$

The right Haar measure is also proportional to the Lebesgue. ■

Friday February 3

■ **Example 1.8** Haan measures for this group:

$$\Omega_\Theta = (0, \infty)$$

$$\mathcal{G} = \{(\theta \mapsto a\theta) : a > 0\}$$

We can show that a group used for distributions like $N(0, \theta)$

Left transformation:

$$L_a = a\theta = \eta$$

if density for θ is $\pi(\theta)$.

$$\begin{aligned}
 \pi_\eta(\eta) &= \pi\left(\frac{\eta}{a}\right) \left| \frac{d\eta/a}{d\eta} \right| \\
 &= \pi\left(\frac{\eta}{a}\right) \frac{1}{a} \\
 \pi_\eta(1) &= \pi\left(\frac{1}{a}\right) \frac{1}{a} \\
 \pi\left(\frac{1}{a}\right) &\propto a \\
 \pi\left(\theta = \frac{1}{a}\right) &\propto \frac{1}{\theta}
 \end{aligned}$$

Right transformation:

$$R_a(\theta) = g_\theta(a) = \theta a = L_a(\theta)$$

Right Haar density is also

$$\pi(\theta) = \frac{1}{\theta}$$

■

■ **Example 1.9** $N(\mu, \sigma^2), \Omega_\Theta = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\}$

Consider group,

$$\mathcal{G} = \{[(\mu, \sigma) \mapsto (c\mu + b, c\sigma) = (g_{b,c}(\mu, \sigma))] : b \in \mathbb{R}, c > 0\}$$

Left transformation:

$$L_{b,c}(\mu, \sigma) = g_{b,c}(\mu, \sigma) = (c\mu + b, c\sigma) = (\tilde{\mu}, \tilde{\sigma})$$

$$\begin{aligned} \pi_\eta(\tilde{\mu}, \tilde{\sigma}) &= \pi\left(\frac{\tilde{\mu} - b}{c}, \frac{\tilde{\sigma}}{c}\right) \left| \frac{\frac{\partial \mu}{\partial \tilde{\mu}}}{\frac{\partial \sigma}{\partial \tilde{\mu}}} \quad \frac{\frac{\partial \mu}{\partial \tilde{\sigma}}}{\frac{\partial \sigma}{\partial \tilde{\sigma}}} \right| \\ &= \pi\left(\frac{\tilde{\mu} - b}{c}, \frac{\tilde{\sigma}}{c}\right) \left| \frac{\frac{1}{c}}{0} \quad 0 \right| \\ &= \pi\left(\frac{\tilde{\mu} - b}{c}, \frac{\tilde{\sigma}}{c}\right) \frac{1}{c^2} \\ \pi_\eta(0, 1) &= \pi\left(\frac{-b}{c}, \frac{1}{c}\right) \frac{1}{c^2} \end{aligned}$$

So if we consider $(\frac{-b}{c}, \frac{1}{c}) = (\mu, \sigma)$ we get that

$$\pi(\mu, \sigma) \propto \frac{1}{\sigma^2}$$

which is the Left Haan measure.

Right transformation:

$$R_{b,c}(\mu, \sigma) = g_{\mu,\sigma}(b, c) = (\mu + \sigma b, c\sigma) = (\tilde{\mu}, \tilde{\sigma}) \neq L_{b,c}$$

$$\begin{aligned} \sigma b + \mu &= \tilde{\mu} \\ \sigma c &= \tilde{\sigma} \\ \sigma &= \frac{\tilde{\sigma}}{c} \\ \mu &= \tilde{\mu} - \sigma b = \tilde{\mu} - \tilde{\sigma} \frac{b}{c} \end{aligned}$$

$$\begin{aligned} \pi_\eta(\tilde{\mu}, \tilde{\sigma}) &= \pi\left(\tilde{\mu} - \tilde{\sigma} \frac{b}{c}, \frac{\tilde{\sigma}}{c}\right) \left| \frac{\frac{\partial \mu}{\partial \tilde{\mu}}}{\frac{\partial \sigma}{\partial \tilde{\mu}}} \quad \frac{\frac{\partial \mu}{\partial \tilde{\sigma}}}{\frac{\partial \sigma}{\partial \tilde{\sigma}}} \right| \\ &= \pi\left(\tilde{\mu} - \tilde{\sigma} \frac{b}{c}, \frac{\tilde{\sigma}}{c}\right) \left| \frac{1}{0} \quad \frac{-b}{\frac{1}{c}} \right| \\ &= \pi\left(\tilde{\mu} - \tilde{\sigma} \frac{b}{c}, \frac{\tilde{\sigma}}{c}\right) \left| \frac{1}{0} \quad \frac{-b}{\frac{1}{c}} \right| \\ &= \pi\left(\tilde{\mu} - \tilde{\sigma} \frac{b}{c}, \frac{\tilde{\sigma}}{c}\right) \frac{1}{c} \end{aligned}$$

If we replace $(\tilde{\mu}, \tilde{\sigma}) = (0, 1)$ we see that

$$\pi(\mu, \sigma) \propto \frac{1}{\sigma}$$

Thus the Right and Left Harr are not the same. ■

1.7.3 Jeffreys' Prior

Haar requires natural group of transformations which is not always available in particular applications. An easily available prior, Jeffreys' prior is constructed using the following principle.

If we assign θ a measure, Π and $\eta = T(\theta)$ is one-to-one transformation, then the prior assigned to η should satisfy $\Pi \circ T^{-1}$

Jeffreys' Prior

Let $f(X|\theta)$ be the likelihood. Let $I(\theta)$ be the Fisher Information

$$I(\theta) = -E \left[\frac{\partial^2}{\partial \theta \partial \theta^T} \log f(X|\theta) | \theta \right]$$

The Jeffreys' Prior is defined as

$$\Pi(\theta) \propto \sqrt{\det I(\theta)}$$

Theorem 1.7.1 — 1.16. Let π_Θ be the Jeffreys' prior density of θ . $\Pi_\Phi(\phi)$ is the Jeffreys' prior density for ϕ where $\phi = h(\theta)$. Here, h is one-to-one. Then we have that

$$\pi_\Phi(\phi) = \pi_\Theta(h^{-1}(\phi)) \left| \det \left(\frac{\partial h^{-1}(\phi)}{\partial \phi} \right) \right|$$

In terms of measure, $\Pi_\Phi = \Pi_\Theta \circ h^{-1}$.

Proof. Let $f_{X|\Phi}(x|\phi)$ represent the likelihood of ϕ . Because Φ and Θ are one-to-one,

$$f_{X|\Phi}(x|\phi) = f_{X|\Theta}(x|\theta) = f_{X|\Theta}(X|h^{-1}(\phi))$$

Conditional distributions only depends on σ -field and two one-to-one variables generate same σ -field.

GET MORE FROM PHOTOS ■

Monday February 6

■ Example 1.10

$$X_1, \dots, X_n \sim N(\theta, \sigma^2)$$

where both θ & σ are unknown.

Find Jeffreys' prior for θ, σ .

$$f(x_i|\theta, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i - \theta)^2}{2\sigma^2}\right\}$$

Table 1.1: Paper Rock Scissor Loss Table

	R	P	S
R	0	-1	1
P	1	0	-1
S	-1	1	0

$$f(x|\theta, \sigma) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left\{-\frac{\sum_i (x_i - \theta)^2}{2\sigma^2}\right\}$$

$$\log(f(x_i|\theta, \sigma)) = -n\sqrt{2\pi\sigma^2} - \frac{(x_i - \theta)^2}{2\sigma^2}$$

$$\frac{\partial^2}{\partial \theta^2} \log(f(x_i|\theta, \sigma)) = -\frac{n}{\theta}$$

$$\frac{\partial^2}{\partial \theta \partial \sigma^2} \log(f(x_i|\theta, \sigma)) = -(\sigma^2)^{-2} \sum (x_i - \theta)$$

$$\frac{\partial^2}{\partial \sigma^2} \log(f(x_i|\theta, \sigma)) = \frac{n}{2}(\sigma^2)^{-2} - (\sigma^2)^{-3} \sum (x_i - \theta)^2$$

Want to take negative expectation of each of theses.

$$I(\theta, \sigma^2) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2}(\sigma^2)^{-2} \end{pmatrix}$$

Want to find the determinant of I. Here it's proportional to σ^{-6} .

$$\sqrt{\det(I)} = \sigma^{-3} = \phi^{\frac{3}{2}}$$

Which gives us our Jeffrey's prior for (θ, σ^2) .

■

1.8 Statistical Decision Theory

Rock, Paper, Scissor.

Player One Action Space: $\Omega_{A,1} = \{R, P, S\}$

Player Two Action Space: $\Omega_{A,2} = \{R, P, S\}$

Loss Function - $L : \Omega_{A,1} \times \Omega_{A,2} \rightarrow \mathbb{R}$

Without loss of generality, we may assume the lose to be player one's loss.

Winner gains one dollar, loser loses a dollar, and in the case of a tie, no money is exchanged.

So we have $(\Omega_{A,1}, \Omega_{A,2}, L)$. In statistics, we may refer to player one as the "statistician" and player two as "nature".

Player one's action is called the action space, Ω_A . Player two's action space is called the parameter space, Ω_Θ .

But where does our data come from in this scenario? We are dealing with a statistical decision problem.

Statistical Decision Problem

The statistician (player one) wants to know nature's action, Θ , but he doesn't. He observes a random variable, whose distribution depends on θ . This is defused information about θ . The random element, X , takes value on

$$(\Omega_X, \mathcal{F}_X)$$

where $X \sim P_\theta$, $\theta \in \Omega_\Theta$.

So we have

$$(\Omega_X, \mathcal{F}_X, \{P_\theta : \theta \in \Omega_\Theta\})$$

on top of $(\Omega_A, \Omega_\Theta, L$.

Decision rule:

$$d : \Omega_X \rightarrow \Omega_A$$

and the collection of all decision rule is \mathcal{D} . This incurs a loss $L(\theta, d(x))$.

Before we take an action, our anticipation of loss is

$$E(L(\theta, d(x)) | \theta)$$

which we may refer to as **risk**.

Game: $(\Omega_A, \Omega_\Theta, L)$

Data: $(\Omega_X, \mathcal{F}_X)$

Model: $\{P_\theta : \theta \in \Omega_\Theta\}$

Decision Rule: \mathcal{D}

Risk: R

Bayesian Statistical Decision Problem

Here, the game, data, model, and decision rule remain the same. The difference is in how we calculate the risk and the use of a prior.

Game: $(\Omega_A, \Omega_\Theta, L)$

Data: $(\Omega_X, \mathcal{F}_X)$

Model: $\{P_\theta : \theta \in \Omega_\Theta\}$

Decision Rule: \mathcal{D}

Prior: P_Θ

Risk: r

$$r(d) = E_{\Theta, X}[L(\Theta, d(x))] = \int_{\Omega_\Theta \times \Omega_X} L(\theta, d(x)) dP_{\Theta, X}$$

Wednesday February 8

Bayes Rule: Optimal estimation, optimal test. This is (mathematically) easier than frequentist optimal procedures because we have a metric in Θ . Don't have to make uniform statements about θ , such as UMVUE, UMP, UMPU,

Definition 1.8.1 — Bayes' Rule. The Bayes rule is

$$d_B = \arg \min \{r(d) : d \in \mathcal{D}\}$$

If P_Θ is improper then d_B is called **generalized Bayes' Rule**.

$$r(d) = \int_{\Omega_\Theta \times \Omega_X} L(\theta, d(x)) f_{X|\Theta}(x, \theta) d(\mu_X \mu_\Theta)(x, \theta)$$

Note that usually $L(\theta, d(x)) \geq 0$ so we have that

$$f_{X|\Theta} = \begin{cases} f(x|\theta)\pi(\theta) \\ \pi(\theta|x)f(x) \end{cases}$$

First way:

$$\begin{aligned} r(d) &= \int_{\Omega_\Theta} \left[\int_{\Omega_X} L(\theta, d(x)) f_{X|\Theta}(x|\theta) d(\mu_X)(x) \right] \pi(\theta) d\mu_\Theta \\ &= \int_{\Omega_\Theta} R(\theta, d) \pi(\theta) d\mu_\Theta \end{aligned}$$

Other way would swap Θ, X . (??)

$$\begin{aligned} r(d) &= \int_{\Omega_X} \left[\int_{\Omega_{\theta}} L(\theta, d(x)) \pi_{\Theta|X}(x|\theta) d(\mu_\Theta)(\theta) \right] d\mu_X \\ &= \int_{\Omega_X} \rho(x, d(x)) d\mu_X(x) \end{aligned}$$

Where we have the posterior expected loss,

$$\rho(x, a) = E(L(\theta, a)|X)$$

INSERT PHOTO

How to calculate Bayes's rule, not by definition.

Definition 1.8.2

$$\arg \min (r(d) : d \in \mathcal{D})$$

is the minimum over a set of functions, \mathcal{D} . A member of \mathcal{D} is

$$d : \Omega_X \rightarrow \Omega_A$$

But this can be converted into minimization over numbers or vectors by the following theorem.

Theorem 1.8.1 — 1.7 in Notes. Suppose $L(\theta, a) \geq C < -\infty$ for all $\theta \in \Omega_\Theta, a \in \Omega_A$. Then the decision rule,

$$d_B : \Omega_X \rightarrow \Omega_A, x \mapsto \arg \min \{\rho(x, a) : a \in \Omega_A\}$$

is the Bayes' Rule.

Ω_A is usually \mathbb{R}, \mathbb{R}^P or subsets thereof. So we are not minimizing over functional spaces.

Proof. By Tonelli's theorem, (because we can assume WLoG that $L(\theta, a) \geq 0, \forall \theta, a$)

$$r(d) = \int_{\Omega_X} \rho(x, d(x)) f_X(x) d\mu_X(x)$$

So, for any x ,

$$\rho(x, d_B(x)) \leq \rho(x, d(x))$$

Thus,

$$\int_{\Omega_X} \rho(x, d_B(x)) f_X(x) d\mu_X(x) \leq \int_{\Omega_X} \rho(x, d(x)) f_X(x) d\mu_X(x)$$

And we have that $r(d_B) \leq r(d) \forall d \in \mathcal{D}$. ■

In frequentist theory, optimality have to be stated uniformly.

"Commonly used optimal criteria" in frequentist decision theory

1. admissibility
2. minimax

Definition 1.8.3 A decision rule is **inadmissible** if there exists $d' \in \mathcal{D}$ such that

1. $R(\theta, d') \leq R(\theta, d) \forall \theta \in \Omega_\Theta$
2. $R(\theta, d') < R(\theta, d)$ for some $\theta \in \Omega_\Theta$

A decision rule that is not inadmissible is admissible.

Definition 1.8.4 A decision rule $d \in \mathcal{D}$ is a **minimax rule** if for all $d' \in \mathcal{D}$,

$$\sup_{\theta \in \Omega_\Theta} R(\theta, d) \leq \sup_{\theta \in \Omega_\Theta} R(\theta, d')$$

The relation between Bayes' rule and admissible rule, generally "all Bayes rule are admissible".

Theorem 1.8.2 — 1.8 in Notes. Suppose

1. for each $d \in \mathcal{D}$, $R(\theta, d)$ is integrable with respect to P_Θ
2. for any $d_1, d_2 \in \mathcal{D}$,

$$R(\theta, d_1) - R(\theta, d_2) < 0$$

for some $\theta \in \Omega_\Theta$ which implies that

$$P(R(\theta, d_1) - R(\theta, d_2) < 0) > 0$$

Then any Bayes or generalized Bayes rule is admissible.

Proof. Suppose $d_1 \in \mathcal{D}$ is Bayes and inadmissible. Then there exists $d_2 \in \mathcal{D}$ such that

1. $R(\theta, d_2) \leq R(\theta, d_1) \forall \theta \in \Omega_\Theta$
2. $R(\theta, d_2) < R(\theta, d_1)$ for some $\theta \in \Omega_\Theta$

But this gives us that

$$R(\theta, d_2) - R(\theta, d_1) \geq 0 \forall \theta$$

and

$$P(R(\theta, d_2) - R(\theta, d_1) < 0) > 0$$

Recall from STAT 517 that if $f \geq 0, \mu(f > 0) > 0$ then

$$\int f d\mu > 0$$

or that if $f \leq 0, \mu(f < 0) > 0$ then

$$\int f d\mu < 0$$

This implies that

$$\begin{aligned} \Rightarrow \int R(\theta, d_2) - R(\theta, d_1) dP_\Theta &< 0 \\ \Rightarrow \int R(\theta, d_2) dP_\Theta &< \int R(\theta, d_1) dP_\Theta \\ &\Rightarrow r(d_2) < r(d_1) \\ &\Rightarrow d_1 \text{ not Bayes' } \end{aligned}$$

■

Friday February 10

Theorem 1.8 condition 2 is a mild condition which is satisfied by the two important cases

1. $R(\theta, d)$ is continuous function for all $d \in \mathcal{D}$, and $P(\Theta \in G) > 0$ for any empty open set.
For example, if $P \ll \lambda, \lambda \ll P (P \equiv \lambda)$, then $P(\Theta \in G) > 0$ for all open $G \neq \emptyset$. To see that the above statement is sufficient for the second condition of Theorem 1.8,
FINISH FROM PHOTO
2. Ω_Θ is countable (σ -finite)

$$\Omega_\Theta = \{\theta_1, \theta_2, \dots\}$$

and

$$P(\Theta = \theta_i) > 0, \forall i = 1, 2, \dots$$

If $R(\theta, d_2) - R(\theta, d_1) > 0$ for some $\theta \in \Omega_\Theta$ then this implies that

$$R(\theta_i, d_2) - R(\theta_i, d_1) > 0 \quad i \in \{1, 2, \dots\}$$

But we know that $P(\Theta = \theta_i) > 0$ so

$$P(R(\theta_i, d_2) - R(\theta_i, d_1) > 0) > 0$$

2. Estimation & Inference

Typically for estimation,

$$\Omega_{\Theta} = \mathbb{R}^P, \Theta \subseteq \mathbb{R}^P, \Omega_A = \Omega_{\Theta}$$

so for testing problems,

$$\Omega_{\Theta} = \{0, 1\} = \Omega_A$$

but for clarification problems,

$$\Omega_{\Theta} = \{a_1, \dots, a_k\} = \Omega_A$$

2.1 Estimation

$$\Omega_{\Theta} = \Omega_A = \mathbb{R}^P$$

One of the most commonly used loss is the squared loss or L_2 loss.

Definition 2.1.1 The L_2 loss is

$$L(\theta, a) = (\theta - a)^T w(\theta)(\theta - a)$$

where $w(\theta) > 0, \forall \theta \in \Omega_{\Theta}$

In this case, the Bayes' rule is explicit.

Theorem 2.1.1 Suppose that

1. $E[\theta^T w(\theta) \theta | X] < \infty$ a.s.
2. $E[w(\theta) | X] > 0$ a.s

Then,

$$d_B(x) = (E[w(\theta) | X])^{-1} E[w(\theta) \theta | X]$$

is the Bayes' rule.

Note that if $w(\theta) = I_p$, then Bayes' rule for loss, $\|\theta - a\|^2$, is simply $E(\theta|X)$.

Proof. Because $L(\theta, a) \geq 0$, we need to minimize

$$\rho(x, a) = E[(\theta - a)^T w(\theta)(\theta - a)|X]$$

where a can depend on x .

$$\begin{aligned} E[(\theta - d_B(x) + d_B(x) - a)^T w(\theta)(\theta - d_B(x) + d_B(x) - a)|X] \\ &= E[(d_B(x) - a)^T w(\theta)(d_B(x) - a)|X] + E[(d_B(x) - a)^T w(\theta)(\theta - d_B(x))|X] + E[(\theta - d_B(x))^T w(\theta)(\theta - d_B(x))|X] \\ &= (d_B - a)^T E[w(\theta - d_B)|X] \\ &= (d_B - a)^T (E[w\theta|X] - E[w d_B|X]) \\ &= 0 \end{aligned}$$

Overview:

$$\rho(x, a) = \text{nonneg} + 0 + 0 + \rho(x, d_B) \Rightarrow \rho(x, a) \geq \rho(x, d_B)$$

So we have that d_B is the Bayes' rule. ■

Another commonly used loss is the L_1 loss.

Definition 2.1.2 If $p = 1$ then the L_1 loss is

$$L(\theta, a) = |\theta - a|$$

What is the Bayes rule here?

Definition 2.1.3 — Median. Let U be a random variable with c.d.f. F . Then any number m that satisfies

$$F(m-) \leq \frac{1}{2} \leq F(m)$$

is called the **median**. Note that this definition does not require F to be monotone.

Theorem 2.1.2 If U is integrable and m is a median of U , then

$$\int |U - m| dP \leq \int |U - a| dP \quad \forall a \in \Omega_U$$

Proof. Case 1. $m < a$

$$\begin{aligned} \int_{\Omega_U} |U - m| - |U - a| dP &= \int_{U \leq m} + \int_{m < U \leq a} + \int_{U < a} \\ &= \int_{U \leq m} m - a dP + \int_{m < U \leq a} 2U - m - a dP + \int_{U < a} a - m dP \\ &= (m - a)P(U \leq m) + (**) + (a - m)P(U > a) \end{aligned}$$

For (**),

$$\int_{m < U \leq a} 2U - m - a dP \leq (2a - m - U)P(m < U \leq a) \quad (2.1)$$

$$= (m - a)[P(U \leq m) - P(U > a) - P(m < U \leq a)] \quad (2.2)$$

$$= (m - a)[P(U \leq m) - P(U > m)] \quad (2.3)$$

$$= (m - a)[P(U \leq m) - (1 - P(U \leq m))] = (m - a)[2F(m) - 1] \quad (2.4)$$

So when $a > m$,

$$E|U - m| - E|U - a| \leq (2F(m) - 1)(m - a)$$

Case 2. $a < m$

By the same argument,

$$E|U - m| - E|U - a| \leq (1 - 2F(m-))(a - m)$$

But, because m is a median,

$$F(m-) \leq \frac{1}{2} \leq F(m)$$

FINISH FROM PHOTO



Monday February 13

Corollary 2.1

The posterior median $M(\theta|X)$ is a Bayes rule with respect to

$$L(\theta, a) = |\theta - a|$$

More generally,

$$L(\theta, a) = \begin{cases} \alpha_1(\theta - a) & \theta \geq a \\ \alpha_2(a - \theta) & \theta < a \end{cases}$$

Here, Bayes rule is quantile.

Besides Bayes estimations, another popular estimation is the generalized MLE. This is simply,

$$\hat{\theta} = \arg \max \{ \pi(\theta|X) : \theta \in \Omega_{\Theta} \}$$

the posterior mode. MLE is a special case where $\pi(\theta)$ is a constant.

■ Example 2.1 Bayes estimation under

$$L(\theta, a) = (\theta - a)^2 = |\theta - a|$$

Also generalized MLE.

We have a linear regression,

$$Y_i = \theta x_i + \varepsilon_i \quad i = 1, \dots, n$$

$$\varepsilon_1, \dots, \varepsilon_n | \theta \stackrel{iid}{\sim} N(\theta, \sigma^2)$$

We have prior knowledge that $\theta \geq 0$ (e.g. growth rate). how do we incorporate their information for estimation.

$$\Pi(\theta) = 1 \quad \theta \geq 0 = I(\theta \geq 0)$$

Here we find the improper prior.

$$f(x|\theta) \propto \exp\left\{-\frac{x^T x}{2\sigma^2}\theta^2 + \frac{x^T y}{\sigma^2}\theta\right\}$$

$$\text{where } x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

By Lemma 1.1, if $f(\theta) \propto \exp\{-a\theta^2 + b\theta\}$ then

$$\theta \sim N\left(\frac{b}{2a}, \frac{1}{2a}\right)$$

So, as a function of θ , $\theta \mapsto f(x|\theta)$ has the form

$$N\left(\frac{x^T y}{x^T x}, \frac{\sigma^2}{x^T x}\right)$$

and we can then see that the posterior density would be

$$\pi(\theta|x) = \frac{N(\mu(x,y), \tau^2(x,y))I(\theta \geq 0)}{\int_0^\infty N(\mu, \tau^2)d\mu}$$

Note that this integral integrated with respect to θ becomes

$$\int_0^\infty N(\mu, \tau^2)d\theta = \int_{\mu/\tau}^\infty N(0, 1)d\theta = \Phi(\mu/\tau)$$

Now with the conjugate Bayes rule and $L(\theta, a) = (\theta - a)^2$ we may let $\gamma = \frac{\theta - \mu}{\tau}$ and get that

$$\pi(\gamma|x) = \frac{N(0, 1)}{\Phi(\frac{\mu}{\tau})} I(\gamma \geq -\frac{\mu}{\tau})$$

$$\begin{aligned} E(\gamma|X) &= \frac{1}{\Phi(\mu/\tau)} \int_{-\mu/\tau}^\infty \gamma N(0, 1) d\gamma \\ &= \frac{\exp\{-\mu^2/(2\tau^2)\}}{\sqrt{2\pi}\Phi(\mu/\tau)} \end{aligned}$$

$$E(\theta|X) = \mu + \tau * \frac{\exp\{-\mu^2/(2\tau^2)\}}{\sqrt{2\pi}\Phi(\mu/\tau)}$$

where, as above, $\mu = \frac{x^T y}{x^T x}$, and $\tau^2 = \frac{\sigma^2}{x^T x}$.

■

Bayes rule under $L(\theta, a) = |\theta - a|$ is the median. So the solution would be

$$\int_0^m N(\mu, \tau^2) d\theta = \frac{1}{2} \int_0^\infty N(\mu, \tau^2) d\theta$$

So with the change of variables,

$$\int_{-\mu/\tau}^{(m-\mu)/\tau} N(0, 1) d\gamma = \frac{1}{2} \int_{-\mu/\tau}^\infty N(0, 1) d\gamma$$

and after some algebra we get that

$$m = \tau \Phi^{-1}\left(1 - \frac{1}{2} \Phi(\mu/\tau)\right) + \mu$$

and so here the generalized MLE is the MLE.

This is simply maximizing $N(\mu, \tau^2)$ over $\theta \geq 0$.

So the generalized MLE over $\theta \geq 0$ is

$$\max\left(0, \frac{x^T y}{x^T x}\right)$$

2.2 Bayes Rule and Unbiasedness

Another relation between Bayesian and Frequentist criteria. Whereas bayes rule and admissibility are somewhat consistent, unbiased is a fundamentally unBayesian idea.

Recall that $d(x)$ is unbiased for θ if and only if

$$E(d(x)|\theta) = \theta \forall \theta$$

Theorem 2.2.1 — 2.3 in notes. Suppose $d_B(x)$ is Bayes rule with respect to

$$L(\theta, a) = (\theta - a)^T W(\theta) (\theta - a)$$

then $d_B(x)$ is biased unless

$$d(x) = \theta \text{ a.s. } P_{\theta, X}$$

Proof. Recall that

$$d_B(x) = [E(W(\theta)|X)]^{-1} E(W(\theta)\theta|X)$$

In the following, write $W(\theta) = W$, $d_B(x) = d_B$. We want to show that if d_B is unbiased then

$$r(d_B) = 0 \Leftrightarrow \theta = d_B(x) \text{ a.s. } P$$

$$\begin{aligned}
r(d_B) &= E[(\theta - E(W|X)^{-1}E(W\theta|X)^T W(\dots))] \\
&= E[E(\theta^T W \theta | X) - E(W\theta|X)^T E(W|X)^{-1}E(W\theta|X)] \\
&= E[E(\theta^T W \theta | X) - d_B^T E(W\theta|X)] \\
&= E[E(\theta^T W \theta - d_B^T W \theta | X)] \\
&= E(\theta^T W \theta - d_B^T W \theta) \\
&= E(\theta^T W \theta - E(d_B|\theta)^T W \theta) \\
&= 0
\end{aligned}$$

■

Wednesday February 15

2.3 Error Assessment in Bayesian Setting

In frequentist setting, X is nothing but one realization of a random variable that potentially takes many values. So we may use

$$\text{Var}_\theta(x) = \text{Var}(x|\theta)$$

or

$$\text{MSE}(X|\theta) = E(X - E(X|\theta))^2$$

In the Bayesian context, X is fixed (i.e. it's being conditional on). Once you condition on something it's out of the probability picture.

$$\begin{aligned}
\text{Var}(\theta|X) &= E[(\theta - E(\theta|X))(\theta - E(\theta|X))^T] \\
\text{MSE}_d(\theta|X) &= E(\theta - d(x)|x)^2 = E[(\theta - d(x))(\theta - d(x))^T]
\end{aligned}$$

Note that $\text{Var}(\theta|X)$ is a special case of $\text{MSE}_d(\theta|X)$ when $d = d_B$ under squared loss. It's easy to show the identity

$$\text{MSE}_d(\theta|X) = \text{Var}(\theta|X) + [E(\theta|X) - d(x)][E(\theta|X) - d(x)]^T$$

■ **Example 2.2** Suppose that

$$X_1, \dots, X_n | \phi \stackrel{iid}{\sim} N(0, \phi), \phi \sim \tau \chi_{(v)}^{-2}$$

Then as in your HW #2, you can show that

$$\phi | X_1, \dots, X_n \sim (s + \tau) \chi_{(v+n)}^{-2}$$

where $s = \sum_{i=1}^n X_i^2$.

Also, as shown in the homework,

$$E(\phi | X_1, \dots, X_n) = \frac{s + \tau}{v + n - 2}$$

$$\text{Var}(\phi | X_1, \dots, X_n) = \frac{2(s + \tau)^2}{(v + n - 2)^2(v + n - 4)}$$

Suppose that $d_1(x) = \frac{1}{n}S$, unbiased estimator ($\mu = 0$). Also that $d_2(x) = \frac{s+\tau}{v+n+2}$ is the GMLE (HW 2).

$$\text{MSE}_{d_1}(\phi|X_1, \dots, X_n) = \frac{2(s+\tau)^2}{(v+n-2)^2(v+n-4)} + \left(\frac{s+\tau}{v+n-2} - \frac{1}{n}S \right)^2$$

$$\text{MSE}_{d_2}(\phi|X_1, \dots, X_n) = \frac{2(s+\tau)^2}{(v+n-2)^2(v+n-4)} + \left(\frac{s+\tau}{v+n-2} - \frac{s+\tau}{v+n+2}S \right)^2$$

So you can present $d \pm \sqrt{\text{MSE}_d(\theta|X_1, \dots, X_n)}$. ■

2.4 Credible Set (or Interval)

In the frequentist setting, CI is a random interval. $C(X)$:

$$P_{\Theta}(\theta \in C(X)) = 1 - \alpha$$

In the Bayesian setting, $C(X)$ is fixed, we want:

$$P(\theta \in C(X)|X) = 1 - \alpha$$

Definition 2.4.1 A $(1 - \alpha)$ -credible set is any $C \in \mathcal{F}_{\theta}$ such that

$$P(\Theta^{-1}(C)||X) \geq 1 - \alpha \quad a.s.$$

Shortest interval is preferred. HPD: Highest Posterior Density Credible Set.

Definition 2.4.2 The $(1 - \alpha)$ -highest posterior density for θ is a $C \in \mathcal{F}_{\theta}$ such that

$$C = \{\theta \in \Omega_{\theta} | \pi(\theta|X) \geq k_{\alpha}\}$$

where

$$k_{\alpha} = \sup\{k | P(\pi(\theta|X) \geq k) \geq 1 - \alpha\}$$

Intuitively,

PHOTO

Continuous $\pi(\theta|X)$ then in this case (PHOTO)

$$P(\pi(\theta|X) \geq k_{\alpha}) = 1 - \alpha$$

$$P(\{\theta | \pi(\theta|X) \geq K_{\alpha}\} | X)$$

As we move K up, $P(\pi(\theta|X) \geq K)$ takes only three values. What if $1 - \alpha$ is not one of the three?

PHOTO

HPD is the shortest credible set as proved in next theorem.

Theorem 2.4.1 — 2.4. Suppose that $\pi(\theta) > 0$, for all $\theta \in \Omega_{\theta}$. Let C_{α}^* be a $(1 - \alpha)$ HPD credible set and C_{α} be any $(1 - \alpha)$ credible set.

Furthermore, assume that $P_{\Theta|X}(C_{\alpha}^*|X) = 1 - \alpha$. Then,

$$\mu_{\Theta}(C_{\alpha}^*) \leq \mu_{\Theta}(C_{\alpha})$$

Proof. Want to show:

$$\mu_{\Theta}(c) < \mu_{\Theta}(C_{\alpha}^*) \Rightarrow P_{\Theta|X}(C|X) < P_{\Theta|X}(C_{\alpha}^*|X) = 1 - \alpha$$

$$\text{Let } C \subseteq \Omega_{\Theta}, \mu_{\Theta}(C) < \mu_{\Theta}(C_{\alpha}^*) \Rightarrow \mu(C \setminus C_{\alpha}^*) < \mu(C_{\alpha}^* \setminus C)$$

Friday February 17

FINSIH PROOF

We know that

$$\pi(\theta|X) \geq K_{\alpha} \text{ on } C_{\alpha}^* \Rightarrow \text{on } C_{\alpha}^* \setminus C.$$

$$\pi(\theta|X) \leq K_{\alpha} \text{ on } (C_{\alpha}^*)^C \Rightarrow \text{on } C \setminus C_{\alpha}^*.$$

$$\begin{aligned} P_{\Theta|X}(C_{\alpha}^* \setminus C|X) &= \int_{C_{\alpha}^* \setminus C} \pi(\theta|x) d\mu_{\Theta}(\theta) \\ &\geq K_{\alpha} \mu_{\Theta}(C_{\alpha}^* \setminus C) \\ &> K_{\alpha} \mu_{\Theta}(C \setminus C_{\alpha}^*) \\ &\geq \int_{C \setminus C_{\alpha}^*} \pi(\theta|X) d\mu_{\Theta} \\ &= P_{\Theta|X}(C \setminus C_{\alpha}^*) \end{aligned}$$

■

INCLUDE PICTURE

■ Example 2.3

$$X_1, \dots, X_n | \lambda, \phi \stackrel{iid}{\sim} N(\lambda, \phi)$$

$$\lambda | \phi \sim N(a, \frac{\phi}{m})$$

$$\phi \sim \tau \chi_{(k)}^2$$

So we know that

$$\frac{\sqrt{n+m}(\lambda - a(x))}{\sqrt{\tau(x) \setminus (n+k)}} | X \sim t_{(n_k)}$$

where,

$$a(x) = (n\bar{x} + ma) \setminus (n+m)$$

$$\tau(x) = \sum (X_i - \bar{X})^2 + \tau + (\bar{X} - a)^2 (m^{-1} + n^{-1})$$

INSERT PHOTO

So the $(1 - \alpha) * 100\%$ credible set for λ is

$$\{\lambda : -t_{(n+k)}(\frac{\alpha}{2}) < \frac{\sqrt{n+m}(\lambda - a(x))}{\sqrt{\tau(x) \setminus (n+k)}} < t_{(n+k)}(\frac{\alpha}{2})\}$$

Here we get,

$$a(x) \pm t_{(n+k)}(\frac{\alpha}{2}) \sqrt{\tau(x) [(n+k)(n+m)]}$$

Recall that we also know that

$$\frac{\phi}{\tau(x)}|x \sim \chi_{(n+k)}^{-2}$$

PHOTO

C_1, C_2 are solution to

$$h(C_1) = h(C_2)$$

$$\int_{C_1}^{C_2} h(t)dt = 1 - \alpha$$

where h is pdf of χ_{k+m}^{-2}

HDD set of $\{\phi : C_1 < \frac{\phi}{\tau(x)} < C_2\}$

■

2.5 Hypothesis Test

PHOTO

Common we use the following losses

- 0-1 <- if wrong lose 1
- 0 - C_1 - C_2 <- more nuanced. If wrong one way, C_1 . If wrong the other way, C_2 .

$$L(\theta, a) = \begin{cases} 0 & (\theta, a) \in (\Omega_{\Theta}^{(0)} x \{a_0\}) \cup (\Omega_{\Theta}^{(1)} x \{a_1\}) \\ C_1 & (\theta, a) \in (\Omega_{\Theta}^{(0)} x \{a_1\}) \\ C_2 & (\theta, a) \in (\Omega_{\Theta}^{(1)} x \{a_0\}) \end{cases}$$

Theorem 2.5.1 Suppose that

$$0 < P_{\Theta|X}(\Omega_{\Theta}^{(0)}|X) < 1$$

then the Bayes' rule for 0 - C_1 - C_2 Loss is

$$d_B(x) = \begin{cases} a_0 & C_1 P_{\Theta|X}(\Omega_{\Theta}^{(1)}|X) \leq C_0 P_{\Theta|X}(\Omega_{\Theta}^{(0)}|X) \\ a_1 & C_1 P_{\Theta|X}(\Omega_{\Theta}^{(1)}|X) > C_0 P_{\Theta|X}(\Omega_{\Theta}^{(0)}|X) \end{cases}$$

Proof. Since the loss is bad, need to minimize

$$\rho(x, a)$$

over (a_0, a_1) .

$$\begin{aligned} \rho(x, a) &= E(L(\theta, a)|X) \\ &= \int_{\Omega_{\Theta}^{(0)}} L(\theta, a) dP_{\Theta|X} + \int_{\Omega_{\Theta}^{(1)}} L(\theta, a) dP_{\Theta|X} \\ \rho(x, a_0) &= 0 + C_1 P_{\Theta|X}(\Omega_{\Theta}^{(1)}) \\ \rho(x, a_1) &= C_2 P_{\Theta|X}(\Omega_{\Theta}^{(0)}) + 0 \end{aligned}$$

■

There is a problem in the frequentist setting (e.g. $H_0 : \theta = \theta_0$)

$$\Omega_{\Theta}^{(0)} = \{\theta_0\}$$

So, if we put an absolute continuous prior, then there is no mass assigned on $\Omega_{\Theta}^{(0)}$. So, the above Bayes rule does not work.

Definition 2.5.1 Let (Ω, \mathcal{F}) be a measureable space. Then the Divas measure for $a \in \Omega$ is the set function

$$\delta_a(B) = \begin{cases} 1 & a \in B \\ 0 & a \notin B \end{cases}$$

This is a measure (proved in 517).

Note that is also can be shown that for any function $f : \Omega \rightarrow \mathbb{R}, B \in \mathcal{F}$

$$\int_B f(\omega) \delta_a(d\omega) = f(a) \delta_a(B)$$

Monday February 20

Lemma 2.1 Suppose that (Ω, \mathcal{F}) is a measureable space. We have that $a \in \Omega, B \in \mathcal{F}, \delta_a(B)$ is the Dirac measure about a. Suppose that for any a, $\{a\} \in \mathcal{F}$, (true for Borel). Suppose that

$$f : \Omega \rightarrow \mathbb{R}, \text{ on } \mathcal{F} \setminus \mathbb{R}$$

then we have that

$$\int_B f(\omega) d\delta_a(\omega) = f(a) \delta_a(B)$$

Proof.

$$\int_B f d\delta_a = \sup_{\{A_i\} \in \mathcal{P}} \sum_{i=1}^k [\inf_{\omega \in A_i} f(\omega)] \delta_a(A_i)$$

If $a \notin B$, then the above summation part of the equality is 0 for all $\{A_i\} \in \mathcal{P}$.

Now suppose that $a \in B$, then there exists a unique $A_i \in \{A_i\}$ such that $a \in A_i$.

So for any $\{A_i\} \in \mathcal{P}$

$$\sum_{j=1}^k [\inf_{\omega \in A_j} f(\omega)] \delta_a(A_j) = [\inf_{\omega \in A_i} f(\omega)] \delta_a(A_i), \quad a \in A_i$$

We know that this term is less than $f(a)$, so this implies that

$$\int_B f d\delta_a \leq f(a)$$

From here we want to take the special partition

$$(\{a\}, B \setminus \{a\}) \in \mathcal{P}$$

$$\begin{aligned}
\int f d\delta_a &\geq [\inf_{\omega \in \{a\}} f(\omega)]\delta_a(\{a\}) + [\inf_{\omega \in B \setminus \{a\}} f(\omega)]\delta_a(B \setminus \{a\}) \\
&= f(a) * 1 + 0 \\
&= f(a)
\end{aligned}$$

So we have that $\int f d\delta_a = f(a)$ and thus we may conclude that

$$\int f d\delta_a = \begin{cases} 0 & a \notin B \\ f(a) & a \in B \end{cases} \Rightarrow \int f d\delta_a = f(a)\delta_a(B)$$

■

Suppose that Q_Θ is a measure on $(\Omega_\Theta, \mathcal{F}_\Theta)$ such that

$$Q_\Theta \ll \nu_\Theta \ll \lambda$$

Let $P_\Theta = (1 - \varepsilon)Q_\Theta + \varepsilon\delta_{\theta_0}$ and $\Pi_\Theta(\theta) = \frac{dQ_\Theta}{d\nu_\Theta}$.

Theorem 2.5.2 — 2.6. Suppose that P_Θ is defined as above. Then, for $a \in A_i$,

$$P_{\Theta|X}(\{\theta_0\}|x) = \frac{\varepsilon f(x|\theta_0)}{(1 - \varepsilon) \int_{\Omega_\Theta} f(x|\theta) \pi_\Theta(\theta) d\nu_\Theta(\theta) + \varepsilon f(x|\theta_0)}$$

Proof. Let $\mu_\Theta = (1 - \varepsilon)\nu_\Theta + \varepsilon\delta_{\theta_0}$. We want to show that $P_\Theta \ll \mu_\Theta$ and find $\frac{dP_\Theta}{d\mu_\Theta}$. But in order to find this derivative we'll have to guess and check!

$$\tau_\Theta(\theta) = \begin{cases} \pi_\Theta(\theta) & \theta \neq \theta_0 \\ 1 & \theta = \theta_0 \end{cases}$$

We want to check $\tau_\Theta = \frac{dP_\Theta}{d\mu_\Theta}$.

Let $B \in \mathcal{F}_\Theta$,

$$\begin{aligned}
\int_B \tau_\Theta(\theta) d\mu_\Theta(\theta) &= \int_B \tau_\Theta(\theta) d((1 - \varepsilon)\nu_\Theta + \varepsilon\delta_{\theta_0}) \\
&= (1 - \varepsilon) \int_B \tau_\Theta d\nu_\Theta + \varepsilon \int \tau_\Theta d\delta_{\theta_0} \\
&= \dots \int_{B \setminus \{\theta_0\}} \tau_\Theta d\nu_\Theta + \dots \\
&= \dots \int_{B \setminus \{\theta_0\}} \Pi_\Theta d\nu_\Theta + \dots \\
&= \dots \int_B \Pi_\Theta d\nu_\Theta + \dots \\
&= (1 - \varepsilon) \int_B \Pi_\Theta d\nu_\Theta + \varepsilon \int \tau_\Theta d\delta_{\theta_0} \\
&= (1 - \varepsilon) \int_B \Pi_\Theta d\nu_\Theta + \varepsilon \delta_{\theta_0}(B) \\
&= (1 - \varepsilon)Q_\Theta(B) + \varepsilon\delta_{\theta_0}(B) \\
&= P_\Theta(B)
\end{aligned}$$

Therefore by RN Theorem

$$P_{\Theta} \ll \mu_{\Theta}, \frac{dP_{\Theta}}{d\mu_{\Theta}}$$

the posterior density from $\tau_{\Theta}(\theta)$ (the prior density) and the likelihood gives us,

$$\tau_{\Theta|X}(\theta|x) = \frac{f_{X|\Theta}(x|\theta)\tau_{\Theta}(\theta)}{\int_{\Omega_{\Theta}} f_{X|\Theta}(x|\theta)\tau_{\Theta}d\mu_{\Theta}(\theta)}$$

So we have that

$$\begin{aligned} P_{\Theta|X}(\{\theta_0\}|x) &= \int_{\{\theta_0\}} \tau_{\Theta|X}(\theta|x)d\mu_{\Theta}(\theta) \\ &= \frac{\int_{\{\theta_0\}} f(x|\theta)\tau_{\Theta}(\theta)d\mu_{\Theta}(\theta)}{\int_{\Omega_{\Theta}} f_{X|\Theta}(x|\theta)\tau_{\Theta}d\mu_{\Theta}(\theta)} \end{aligned}$$

By Lemma 2.1,

$$\begin{aligned} \int_{\{\theta_0\}} f(x|\theta)\tau_{\Theta}(\theta)d\mu_{\Theta}(\theta) &= (1-\varepsilon) \int_{\{\theta_0\}} \dots d\nu_{\Theta}(\theta) + \varepsilon \int_{\{\theta_0\}} \dots d\delta_{\theta_0} \\ &= 0 + \varepsilon f(x|\theta_0) \end{aligned}$$

$$\int_{\Omega_{\Theta}} f_{X|\Theta}(x|\theta)\tau_{\Theta}d\mu_{\Theta}(\theta) = (1-\varepsilon) \int f(x|\theta)\pi(\theta)d\mu(\theta) + \varepsilon \int f(x|\theta)\pi(\theta)d\delta_{\theta_0}(\theta)$$

Therefore,

$$P_{\Theta|X}(\{\theta_0\}|x) = \frac{\varepsilon f(x|\theta_0)}{(1-\varepsilon) \int f(x|\theta)\pi(\theta)d\nu(\theta) + \varepsilon f(x|\theta_0)}$$

■

Corollary 2.2 Consider setting

$$H_0 : \theta = \theta_0, H_1 : \theta \neq \theta_0$$

and action $\mathcal{A} = \{a_0, a_1\}$ where we accept and fail to reject H_0 , respectively.

Suppose the loss function is

$$L(\theta, a) = \begin{cases} 0 & (\theta, a) \in (\{\theta_0\} \times \{a_0\}) \cup (\{\theta_0\}^C \times \{a_1\}) \\ C_0 & (\theta, a) \in (\{\theta_0\} \times \{a_1\}) \\ C_1 & (\theta, a) \in (\{\theta_0\}^C \times \{a_0\}) \end{cases}$$

and the prior is

$$P_{\Theta} = (1-\varepsilon)Q_{\Theta} + \varepsilon\delta_{\theta_0}$$

the Bayes' rule is

$$\frac{\varepsilon f(x|\theta_0)}{(1-\varepsilon) \int f(x|\theta)\pi(\theta)d\nu(\theta) + \varepsilon f(x|\theta_0)} < \frac{C_1}{C_0 + C_1}$$

■ **Example 2.4** Suppose that

$$X_1, \dots, X_n | \theta \stackrel{iid}{\sim} \text{Exp}(\theta)$$

Then the likelihood

$$f(x|\theta) = \theta^{-n} e^{-t(x)/\theta}, \quad \theta > 0$$

$$t(x) = \sum_{i=1}^n X_i$$

Suppose that $\theta \sim \tau \chi_{(m)}^{-2}$. Then (you may check)

$$\theta | \underline{X} \sim (2t(x) + \tau) \chi_{(m+2n)}^{-2}$$

Suppose we want to test

$$H_0 : \theta \leq a, H_1 : \theta > a$$

The Bayes' rule: reject if

$$P(\Omega_{\Theta}^{(1)} | x) > \frac{C_0}{C_0 + C_1}$$

Note that

$$\begin{aligned} P(\Omega_{\Theta}^{(1)} | x) &= P(\theta > a | x) \\ &= 1 - P(\theta \leq a | x) \\ &= 1 - P\left(\frac{\theta}{2t(x) + \tau} \leq \frac{a}{2t(x) + \tau} | x\right) \\ &= 1 - F\left(\frac{a}{2t(x) + \tau}\right) \end{aligned}$$

where F is the c.d.f. of $\chi_{(2n+m)}^{-2}$.

We want to solve

$$a = (2t(x) + \tau) F^{-1}\left(\frac{C_1}{C_0 + C_1}\right)$$

■

Wednesday February 22

Suppose we want to test

$$H_0 : \theta = \theta_0, H_1 : \theta \neq \theta_0$$

Under the $0 - C_0 - C_1$ loss and the slab & spike prior.

The Bayes' rule: reject if

$$\frac{\varepsilon f(x|\theta_0)}{(1 - \varepsilon) \int_{\Omega_{\theta_0}} f(x|\theta) \pi(\theta) d\nu(\theta) + \varepsilon f(x|\theta_0)} < \frac{C_1}{C_0 + C_1}$$

$$f(x|\theta_0) = \theta_0^{-1} e^{-t(x)/\theta_0}$$

$$\pi(\theta) = \tau \chi_{(\nu)}^{-2} \quad \text{slab}$$

$$\pi(\theta) = \frac{1}{\Gamma(\frac{\nu}{2}) 2^{\frac{\nu}{2}}} \left(\frac{\theta}{\tau}\right)^{-\frac{\nu}{2}-1} e^{-\frac{\tau}{2\theta}} \tau^{-1}$$

Take the integral from the rejection denominator, and using the information above

$$\begin{aligned} \int_{-\infty}^{\infty} \theta^{-n} e^{-t(x)/\theta} \frac{1}{\Gamma(\frac{\nu}{2}) 2^{\frac{\nu}{2}}} \left(\frac{\theta}{\tau}\right)^{-\frac{\nu}{2}-1} e^{-\frac{\tau}{2\theta}} \tau^{-1} d\theta &= \text{Algebra to isolate new pdf (integrates to 1)} \\ &= \frac{\Gamma(\frac{2n+2}{2})}{\Gamma(\frac{\nu}{2})} \tau^{\frac{\nu}{2}} (2t(x) + \tau)^{\frac{2n+\nu}{2}} 2^{-n} \end{aligned}$$

2.6 Classification

Ω_{Θ} is a finite set true labels of classes.

$$\Omega_{\Theta} = \{1, \dots, k\}$$

$$\Omega_A = \{1, \dots, k\}$$

PHOTO

The loss function,

$$L: \{1, \dots, k\} \times \{1, \dots, k\}$$

PHOTO

The 0-1 loss is a special case in this.

PHOTO

The 0-1 loss is also a further special case where $C_{01} = C_{10} = 1$

Theorem 2.6.1 The Bayes' Rule for $(\Omega_{\Theta}, \Omega_A, L)$ is

$$d_B(x) = \arg \min \left\{ \sum_{\theta=1}^k C_{\theta a} f_{x|\theta}(x|\theta) \pi_{\Theta}(\theta) : a = 1, \dots, k \right\}$$

Proof. Because the loss function we need to minimize,

$$\rho(x, a) : a = 1, \dots, k$$

Recall that,

$$\begin{aligned} \rho(x, a) &= E(L(\theta, a) | X) \\ &= \sum_{\theta=1}^k L(\theta, a) \pi(\theta | x) \\ &= \sum_{\theta=1}^k C_{\theta a} \pi(\theta | x) \\ &= \sum_{\theta=1}^k C_{\theta a} f(x|\theta) \pi(\theta) \end{aligned}$$

Above, note that

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$$

■

usually, there is a training set and a testing test. In the training set, we know the 'labels', whereas in the testing set we do not.

We have for label θ ,

$$X_{\theta 1}, \dots, X_{\theta n_\theta} \stackrel{iid}{\sim} P_{X|\theta}$$

where $\theta = 1, \dots, k$. An example would be handwriting for letters/numbers.

Generally, there are extra parameters Ψ_θ that determines the shape of clan θ .

$$X_{\theta 1}, \dots, X_{\theta n_\theta} \sim P_{X|\theta, \Psi_\theta}$$

Most commonly used models for $P_{X|\theta, \Psi_\theta}$ is

1. LDA - $N(\mu_\theta, \Sigma)$
2. QDA - $N(\mu_\theta, \Sigma_\theta)$

The parameter, Φ_θ is estimated from training sets, whereas $\pi(\theta)$ is estimated by

$$\frac{n_\theta}{\sum_{\theta=1}^k n_\theta}$$

Plug these into the Bayes then you have LDA, QDA. So, under first model if we use UMVUE to estimate μ_θ, Σ , then

$$\hat{\mu}_\theta = \frac{1}{n_\theta} \sum_{i=1}^{n_\theta} X_{\theta i}$$

$$\hat{\Sigma} = \frac{1}{\sum_{\theta=1}^k n_\theta - k} \sum_{\theta=1}^k \sum_{i=1}^{n_\theta} (x_{\theta i} - \hat{\mu}_\theta)(x_{\theta i} - \hat{\mu}_\theta)^T$$

Plug in Bayes rule and we get LDA.

For the second model,

$$\hat{\mu}_\theta = \frac{1}{n_\theta} \sum_{i=1}^{n_\theta} X_{\theta i}$$

$$\hat{\Sigma} = \frac{1}{n_\theta - 1} \sum_{\theta=1}^k \sum_{i=1}^{n_\theta} (x_{\theta i} - \hat{\mu}_\theta)(x_{\theta i} - \hat{\mu}_\theta)^T$$

Plug this into Bayes rule and get QDA.

Suppose we make a plot using QDA,

PHOTO

2.7 Stein's Estimate

Stein's Estimate

Stein (1956) shows that if the dimension is greater than 3 and sample size is 1, then MLE is not admissible.

Friday February 24

Lemma 2.2 Suppose that

$$X \sim N(\mu, \sigma^2), g: \mathbb{R} \rightarrow \mathbb{R}$$

where g is differentiable,

$$E|g(x)| < \infty$$

Then we have that

$$\text{Cov}(X, g(X)) = \text{Var}(x) E g'(x)$$

Proof. First we assume that

$$X \sim N(0, 1)$$

Let ϕ be the pdf of $N(0, 1)$. Recall that

$$\phi(x) = -x\phi'(x)$$

$$E[g'(x)] = \int_{-\infty}^{\infty} g'(x)\phi(x)dx$$

Let $a \in \mathbb{R}$, then this integral above becomes

$$\begin{aligned} \int_{-\infty}^a g'(x)\phi(x)dx + \int_a^{\infty} g'(x)\phi(x)dx &= \int_{-\infty}^a g'(x) \left(\int_{-\infty}^x \phi'(z)dz \right) dx + \int_a^{\infty} g'(x) \left(\int_x^{\infty} \phi'(z)dz \right) dx \\ &= \int_{-\infty}^a \int_{-\infty}^x g'(x)\phi'(z)dzdx + \int_a^{\infty} \int_x^{\infty} g'(x)\phi'(z)dzdx \\ &= \int_{-\infty}^a \int_z^a g'(x)\phi'(z)dx dz + \int_a^{\infty} \int_z^a g'(x)\phi'(z)dx dz \\ &= \int_{-\infty}^a \phi'(z)(g(a) - g(z))dz + \int_a^{\infty} \phi'(z)((g(z) - g(z)))dz \\ &= \int_{-\infty}^a \phi'(z)(g(a) - g(z))dz \\ &= g(a) \int_{-\infty}^a \phi'(z)dz - \int_{-\infty}^a \phi'(z)g(z)dz \\ &= \int_{-\infty}^a zg'(z)\phi(z)dz \\ &= E(Xg(X)) \\ &= \text{Cov}(X, g(X)) \\ &= E(g'(X)) \end{aligned}$$

More generally, if $X \sim N(\mu, \sigma^2)$ then $X = \sigma Z + \mu$ where $Z \sim N(0, 1)$.

$$\text{Cov}(X, g(X)) = \sigma \text{Cov}(Z, g_1(z))$$

Note that

$$g_1(z) = g(\sigma z + \mu) \quad (2.5)$$

$$= \sigma E(g_1'(z)) \quad (2.6)$$

$$= \sigma E(\sigma g_1'(\sigma Z + \mu)) \quad (2.7)$$

$$= \sigma^2 E(g_1'(\sigma Z + \mu)) \quad (2.8)$$

$$= \sigma^2 E(g_1'(x)) \quad (2.9)$$

$$= \text{Var}(x) E g_1'(x) \quad (2.10)$$

■

Lemma 2.3 (HW #4). If

$$X \sim N(\mu, \Sigma), g : \mathbb{R}^P \rightarrow \mathbb{R}^P$$

where g is differentiable such that $\frac{\partial g}{\partial x^T}$ has integrable components. Then

$$\text{Cov}(X, g(X)) = \Sigma E \left(\frac{\partial g^T(X)}{\partial X} \right)$$

Theorem 2.7.1 — 2.8. Suppose that

$$X \sim N(\theta, I_P)$$

where $P \geq 3$.

Let

$$L(\theta, a) = \|\theta - a\|^2$$

and $d(x) = X$. Then we have that d is inadmissible. ("Very weird")

Proof. Let $d_1(x) = (1 - g(X))X$ where $h : \mathbb{R}^P \rightarrow \mathbb{R}$ is to be specified later.

$$\begin{aligned} R(\theta, d_1) &= E\|\theta - d_1(X)\|^2 \\ &= E\|X - \theta - h(X)X\|^2 \\ &= E\|X - \theta\|^2 - 2E[(X - \theta)^T h(X)X] + E[h^2(X)\|X\|] \\ &= E\|X - \theta\|^2 - 2\text{tr}[\text{Cov}(X, h(X)X)] + E[h^2(X)\|X\|] \\ &= E\|X - \theta\|^2 - 2\text{tr}[I_P E \frac{\partial h(X)}{\partial X}] + E[h^2(X)\|X\|] \\ &= E\|X - \theta\|^2 - 2E(X^T (\frac{\partial h}{\partial X} + ph(X))) + E[h^2(X)\|X\|] \\ &= R(\theta, d) - 2E(X^T (\frac{\partial h}{\partial X} + ph(X))) + E[h^2(x)\|X\|^2] \end{aligned}$$

Let $h(x) = \frac{\alpha}{\|x\|^2}$. Then

$$\frac{\partial h}{\partial X} = \alpha \frac{\partial}{\partial X} (||X||^2)^{-1} \quad (2.11)$$

$$= \alpha(-1)(||X||^2)^{-2} \frac{dX^T X}{dX} = -\alpha ||X||^{-4} 2x \quad (2.12)$$

$$= -\alpha 2X ||X||^{-2} \quad (2.13)$$

So if we plug this into equations above we get

$$||X||^{-2} \alpha (4 - 2P + \alpha)$$

So we have that

$$R(\theta, d_1) - R(\theta, d) = \alpha(4 - 2P + \alpha)E||X||^2$$

If $4 - 2P + \alpha < 0$ then d is inadmissible.

Any $0 < \alpha < 2P - 4$ makes this happen. Note we need $P \geq 3$. ■

Monday February 27

We've said that

$$(1 - \frac{\alpha}{||X||^2})X$$

is better than X for $0 < \alpha < 2p - 4, p \geq 3$. So if we take $\alpha = p - 2$ then we have

$$(1 - \frac{p-2}{||X||^2})X$$

which is called the James-Stein estimator.

2.8 Empirical Bayes

Sometimes we have parameters in Π such as

$$X|\mu \sim N(\mu, 1)$$

$$\mu \sim N(a, \phi)$$

Where before we treated a and ϕ as known hyperparameters.

For Empirical Bayes these parameters will also be estimated, $\pi(\theta|b)$.

The likelihood is

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta|b)$$

We can integrate

$$f(x|b) = \int f(x|\theta)\pi(\theta|b)d\mu_\theta$$

and treat this as our likelihood with b as the parameter. So then you can estimate b by frequentist methods, such as MLE or UMVUE. Once you have \hat{b} you can plug it into the prior, $\pi(\theta|\hat{b})$. With the 'pretend' prior we may carry out Bayesian analysis, such as Bayes rule, generalized MLE, etc. This process is called Empirical Bayes.

■ **Example 2.5** Let $\theta = \theta_1, \dots, \theta_P$. And that $X_1, \dots, X_P | \theta$. Assume that

1. $\theta_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp \theta_P$
2. $X_i | \theta_1, \dots, \theta_P \sim N(\theta_i, 1)$

This condition also implies that

$$X_i \perp\!\!\!\perp \theta^{-i} | \theta^i$$

3. $\theta_i \sim N(0, b)$

Here the likelihood is,

$$f(\underline{X} | \underline{\theta}) = f(X_1 | \underline{\theta}) \cdots f(X_P | \underline{\theta}) = f(X_1 | \theta_1) \cdots f(X_P | \theta_P)$$

And our prior becomes

$$\pi(\underline{\theta} | b) = \pi(\theta_1 | b) \cdots \pi(\theta_P | b)$$

We may integrate to get $f(x | \theta)$

$$\int f(x | \theta) \pi(\theta | b) d\theta = \int \cdots \int f(x_1 | \theta_1) \cdots f(X_P | \theta_P) \pi(\theta_1 | b) \cdots \pi(\theta_P | b) d\theta \quad (2.14)$$

$$= \int f(X_1 | \theta_1) \pi(\theta_1 | b) d\theta_1 \cdots \int f(X_P | \theta_P) \pi(\theta_P | b) d\theta_P \quad (2.15)$$

$$= f(X_1 | b) \cdots f(X_P | b) \quad (2.16)$$

We will derive UMVUE for b . Before that, we first will do Bayes rule for θ under $\|\theta - a\|^2$.

$$d_B(\theta) = E(\theta | X) = \begin{pmatrix} E(\theta_1 | X_1, \dots, X_P) \\ \vdots \\ E(\theta_P | X_1, \dots, X_P) \end{pmatrix}$$

Where

$$E(\theta_i | X_1, \dots, X_P) = \int \theta_i \pi(\theta_i | X_1, \dots, X_P) d\theta_i$$

$$\begin{aligned} \pi(\theta_i | \underline{X}) &= \frac{f(\underline{X} | \theta_i) \pi(\theta_i)}{f(\underline{X})} \\ &= \frac{f(X_1 | \theta_i) \cdots f(X_P | \theta_i) \pi(\theta_i)}{f(\underline{X})} \end{aligned}$$

Note that conditional independence satisfies 4 axioms (see below). ■

Axioms of Conditional Independence

$$X \perp\!\!\!\perp Y | Z$$

Visualized as:

$$\textcircled{X} - \textcircled{Z} - \textcircled{Y}$$

Where X only can "communicate" with Y through Z.

1. $X \perp\!\!\!\perp Y|Z \Rightarrow Y \perp\!\!\!\perp X|Z$
2. $X \perp\!\!\!\perp (Y, Z)|W \Rightarrow X \perp\!\!\!\perp Y|W$
3. $X \perp\!\!\!\perp (Y, Z)|W \Rightarrow X \perp\!\!\!\perp Z|(W, Y)$
4. $X \perp\!\!\!\perp W|(Y, Z) \& X \perp\!\!\!\perp Y|Z \Rightarrow X \perp\!\!\!\perp (Y, W)|Z$

Wednesday March 1

MIDTERM EXAM - Content up to (and including) Section 2.2. Note: 2016 Midterm covered more content.

Friday March 3

■ **Example 2.6**

$$X_1, \dots, X_P | \theta_1, \dots, \theta_P \sim N(\theta_i, 1)$$

$$\theta_i \sim N(0, b)$$

where,

$$\theta_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp \theta_P$$

and b is known.

Bayes rule: $E(\theta_i | \underline{X})$

So we know that

$$\begin{aligned} \pi(\underline{\theta} | \underline{X}) &= \frac{f(\underline{X} | \underline{\theta}) \pi(\underline{\theta})}{f(\underline{X})} \\ &= \frac{f(X_1 | \theta_1) \pi(\theta_1)}{f(X_1)} \dots \frac{f(X_P | \theta_P) \pi(\theta_P)}{f(X_P)} \end{aligned}$$

■

Now we can find,

$$\begin{aligned} E(\theta_i | \underline{X}) &= \int \dots \int \theta_i \pi(\underline{\theta} | \underline{X}) d\underline{\theta} \\ &= \int \dots \int \theta_i \pi(\theta_1 | X_1) \dots \pi(\theta_P | X_P) d\underline{\theta} \\ &= \int \theta_i \pi(\theta_i | X_i) d\theta_i \\ &= E(\theta_i | X_i) \end{aligned}$$

But,

$$X_i | \theta_i \sim N(\theta_i, 1)$$

$$\theta_i \sim N(0, b)$$

$$\theta_i | X_i \sim N\left(\frac{X_i}{\frac{1}{b} + 1}, \frac{1}{\frac{1}{b} + 1}\right)$$

Using Bayes' Rule,

$$E(\theta_i|X_i) = \frac{X_i}{\frac{1}{b} + 1} = (1 - \frac{1}{1+b})X_i$$

and if we know b ,

$$d_B(x) = \begin{pmatrix} \frac{X_1}{\frac{1}{b}+1} \\ \vdots \\ \frac{X_p}{\frac{1}{b}+1} \end{pmatrix}$$

We want to estimate b using UMVUE of $\frac{1}{b+1}$.

Note that $X_i \sim N(0, b+1)$. So, a complete, sufficient statistic would be

$$S = \sum_{i=1}^n X_i^2$$

We know that

$$\frac{S}{b+1} \sim \chi_p^2$$

and the inverse,

$$\frac{b+1}{S} \sim \chi^{-2_p}$$

where we know the expectation is $\frac{1}{p-2}$, rearranging we get

$$E\left(\frac{p-2}{S}\right) = \frac{1}{b+1}$$

giving us the UMVUE of $\frac{1}{b+1}$. Plugging this into Bayes' rule,

$$d_B(x) = \left(1 - \frac{p-2}{||X||^2}\right)X$$

Note that this is the James Stein estimate.

3. Testing Hypothesis for Simple-VS-Simple

3.1 Basic Concepts

(Ω, \mathcal{F}) is measurable space.

(Ω, \mathcal{F}_X) is measurable space.

$X : \Omega \rightarrow \Omega_X \subseteq \mathcal{F} \setminus \mathcal{F}_X$

Let $\mathcal{P}_0, \mathcal{P}_1$ be two disjoint families of probability measure on (Ω, \mathcal{F}) .

P is the true distribution. Assume that $P \in \mathcal{P}_0 \cup \mathcal{P}_1 = \mathcal{P}$

Hypothesis Test

$$H_0 : P \in \mathcal{P}_0, H_1 : P \in \mathcal{P}_1$$

We want to decide whether to reject H_0 .

If everything is deterministic,

$$A : \text{Theory} \Rightarrow B_{Obs}$$

So if B doesn't happen, then A is false.

But in non-deterministic world,

$$A : \text{True} \Rightarrow P(B \text{ occurs}) = \text{large}$$

$$A : \text{Prob not True} \Leftarrow B \text{ doesn't happen}$$

Statistical Inference!

We reject H_0 if X falls in a certain region, $C \subseteq \Omega_X$. This region of rejection is called the...rejection region.

Definition 3.1.1 Let

$$\alpha = \sup_{P \in \mathcal{P}_0} P(X \in C)$$

be the Type I error.

$$\beta(P) = P(X \in C), P \in \mathcal{P}_1$$

is the Power at P. Note that

$$1 - \beta(P)$$

is the Type II error.

If we introduce,

$$\phi : \Omega_X \rightarrow \{0, 1\}$$

$$\phi(x) = \mathbb{I}_{\{C\}}(X)$$

then $P(X \in C) = E(\phi(x))$.

So, more generally, we don't have to restrict range of ϕ to 0, 1.

$$\phi : \Omega_X \rightarrow [0, 1]$$

So, $\phi(x)$ is the probability of rejection given X.

In this general setting the function

Monday March 13

3.2 Neyman-Pearson Lemma

Scenario: You are in a boat arriving at Treasure Island. Your boat can carry 100 pounds of bounty, but no more. On the island there are 1 lb units of gold, diamonds, watermelon, and bananas. Clearly, with each unit being equal in weight you want to take the most valuable goods. Likely diamonds, then gold, then the food. If they are not equal weight you still want to determine which is the most valuable per unit weight.

This is the essence of the Neyman-Pearson Lemma with analogies to likelihood and rejection regions.

Weight is analogous to null probability.

Value is to power.

Weight limit is to alpha.

When we test simple versus simple hypothesis,

$$H_0 : P = P_0 \quad H_1 : P = P_1$$

where $\mathcal{P}_0 = \{P_0\}$, $\mathcal{P}_1 = \{P_1\}$.

Definition 3.2.1 A test ϕ for the simple versus simple hypothesis is Most Powerful size α test (MP- α) if

1. $E_{P_0} \phi(x) = \alpha$
2. If ϕ' is level α . (i.e $E_{P_0} \phi' \leq \alpha$)
then $E_{P_1} \phi \geq E_{P_1} \phi'$

We can assume that P_0, P_1 are dominated by a measure μ . For example, let

$$\mu = P_1 + P_2$$

then, $P_1 \ll \mu, P_2 \ll \mu$.

Lemma 3.1 (Neyman-Pearson: Optimality) Let f_0, f be $\frac{dP_0}{d\mu}, \frac{dP_1}{d\mu}$. Then for any test of the form:

$$\phi(x) = \begin{cases} 1 & f_1(x) > kf_0(x) \\ 0 & f_1(x) < kf_0(x) \end{cases}$$

where $K \geq 0$ is MP- α where $\alpha = E_{P_0}\phi$.

Proof. Let ϕ' be any test of level $E_{P_0}\phi$. We want to show that

$$E_{P_1}\phi \geq E_{P_1}\phi'$$

Consider the formula,

$$(\phi - \phi')(f_1 - kf_0)$$

and look at which term is positive or negative and note that together the terms are always positive.

So,

$$\begin{aligned} \int (\phi - \phi')(f_1 - kf_0)d\mu &> 0 \\ \int (\phi - \phi')f_1d\mu &\geq k \int (\phi - \phi')f_0d\mu \\ \int (\phi - \phi')f_1d\mu &\geq kE_{P_0}\phi - E_{P_0}\phi' \\ E_{P_0}\phi &\geq E_{P_0}\phi' \end{aligned}$$

■

Existence & Uniqueness Part of N-P Lemma

Proposition 3.1 Suppose that $\rho : \mathbb{R} \rightarrow \mathbb{R}$ is right continuous, nonincreasing such that

1. $\rho(0-) = 1$
2. $\lim_{x \rightarrow \infty} \rho(x) = 0$

Then for any $0 < \alpha < 1$ there exists $k \in [0, \infty]$ such that $\alpha \in [\rho(k), \rho(k-)]$.

We may think of this as a discontinuous version of intermediate value theorem.

Proof. If $\alpha \equiv 1$ then we may take $k = 0$ and then

$$\rho(0-) = 1 \geq \rho(0)$$

Now we may assume that $0 < \alpha < 1$. Let $S_\alpha = \{x : \rho(x) \leq \alpha\}$. Let $k = \inf S_\alpha$. Now, because $\alpha > 0, \rho(x) \rightarrow 0, S_\alpha \neq \emptyset$ for any $1 > \alpha > 0, 0 < \inf S_\alpha < \infty$. Let $\{x_n\} \subset S_\alpha$ such that $x_n \downarrow k$ by right continuity, $\rho(x_n) \rightarrow \rho(k)$. Since we have that $\rho(k) \geq \alpha$ for all n, this implies that $\rho(k) \leq \alpha$.

Since $\rho(x) < \alpha$ for all $x < k$, then $\rho(k-) \geq \alpha$. ■

Theorem 3.2.1 For any $0 < \alpha \leq 1$ there exists a test ϕ of size α for the NPL form, i.e.,

$$\phi(x) = \begin{cases} 1 & f_1 > kf_0 \\ 0 & f_1 < kf_0 \end{cases}$$

such that $\phi(x) \equiv \gamma$ for $f_1 = kf_0$. For some $\gamma \in [0, 1], 0 \leq k < \infty$ (the existence part of NPL). Further more, if ϕ' is MP- α then it has to have the form above (almost surely P_0, P_1).

Wednesday March 15

Proof. Existence. For a test of the form:

$$\phi(x) = \begin{cases} 1 & f_1 > kf_0 \\ \gamma & f_1 = kf_0 \\ 0 & f_1 < kf_0 \end{cases}$$

$$\begin{aligned} E_{P_0} \phi(x) &= \int_{f_1 > kf_0} \phi f_0 d\mu + \int_{f_1 = kf_0} \phi f_0 d\mu + \int_{f_1 < kf_0} \phi f_0 d\mu \\ &= P_0(f_1 > kf_0) + \gamma P_0(f_1 = kf_0) + 0 \end{aligned}$$

Let

$$\begin{aligned} \rho(k) &= P_0(f_1 > kf_0) \\ &= P_0(f_1 > kf_0, f_0 > 0) \\ &= P_0\left(\frac{f_1}{f_0} > k, f_0 > 0\right) \end{aligned}$$

We claim that $\rho(k)$ satisfies the four conditions in Proposition 3.1

- Right Continuous

Let $k_n \downarrow k$,

$$\bigcup_{n=1}^{\infty} \{f_1 > k_n f_0\} = \{f_1 > k f_0\}$$

And by continuity of probability,

$$P_0(f_1 > k_n f_0) \rightarrow P_0(f_1 > k f_0)$$

- Nonincreasing
- $\rho(0-) = 1$
Let $k_n \uparrow 0$. Then

$$\bigcap_{n=1}^{\infty} \{f_1 > k_n f_0\} = \{f_1 > 0\}$$

And by continuity of probability,

$$P_0(f_1 > k_n f_0) \rightarrow P_0(f_1 > 0)$$

So,

$$\rho(0-) = 1$$

- $\lim_{k \rightarrow \infty} \rho(k) = 0$
Let $k_n \rightarrow \infty$. Then

$$\bigcap_{n=1}^{\infty} \{f_1 > k_n f_0\} = \emptyset$$

And by continuity of probability,

$$P_0(f_1 > k_n f_0) \rightarrow P_0(\emptyset)$$

So,

$$\rho(k_n) \rightarrow 0$$

So by Proposition 3.1, there exists $0 \leq k_0 < \infty$ such that

$$\rho(k_0) \leq \alpha \leq \rho(k_0-)$$

which is to say,

$$P_0(f_1 > k f_0) \leq \alpha \leq P_0(f_1 \geq k_0 f_0)$$

Let

$$\gamma = \begin{cases} \frac{\alpha - P_0(f_1 > k_0 f_0)}{P_0(f_1 = k_0 f_0)} & P(f_1 = k_0 f_0) = 0 \\ 0 & o.w. \end{cases}$$

Now let

$$\phi(x) = \begin{cases} 1 & f_1 > k f_0 \\ \gamma & f_1 = k f_0 \\ 0 & f_1 < k f_0 \end{cases}$$

If $P_0(f_1 = k f_0) > 0$ then,

$$\begin{aligned} \int \phi f_0 d\mu &= P_0(f_1 > k f_0) + \frac{\alpha - P_0(f_1 > k_0 f_0)}{P_0(f_1 = k_0 f_0)} \\ &= \alpha \end{aligned}$$

And if $P_0(f_1 = k f_0) = 0$ then, $\gamma = 0$ and thus,

$$\begin{aligned} \int \phi f_0 d\mu &= P_0(f_1 > k f_0) \\ &= P_0(f_1 \geq k f_0) \\ &= \alpha \end{aligned}$$

And now for the uniqueness part.

First let us understand what is the meaning of ϕ' with the form,

$$\phi'(x) = \begin{cases} 1 & f_1 > k f_0 \\ 0 & f_1 < k f_0 \end{cases}$$

almost surely P_0, P_1 .

This means that $\phi' = \phi$ on $f_1 \neq k f_0$ almost surely. Then

$$P_{0,1}(\{\omega : f_1(\omega) \neq kf_0(\omega) \rightarrow \phi'(\omega) = \phi(\omega)\}) = 1$$

$$P_{0,1}(\{\omega : f_1(\omega) \neq kf_0(\omega) \rightarrow \phi'(\omega) \neq \phi(\omega)\}) = 0$$

So we want to show that

$$P_{0,1}(f_1 \neq kf_0, \phi' \neq \phi) = 0$$

Since ϕ' is another MP- α test

$$\int \phi' f_0 d\mu = \int \phi f_0 d\mu = \alpha$$

Because both are MP,
PHOTO

■

3.3 Uniformly Most Powerful Test for One-Sided Hypothesis

3.3.1 General Definition

$(\Omega, \mathcal{F}, \mathcal{P})$ is our probability measure where

$$\mathcal{P} = \mathcal{P}_0 \cup \mathcal{P}_1, \mathcal{P}_0 \cap \mathcal{P}_1 = \emptyset$$

We hypothesis that

$$H_0 : P \in \mathcal{P}_0 \quad H_1 : P \in \mathcal{P}_1$$

Definition 3.3.1 A test ϕ for

$$H_0 : P \in \mathcal{P}_0 \quad H_1 : P \in \mathcal{P}_1$$

is UMP of size α (UMP- α) if

1. it is of size α

$$\sup_{P \in \mathcal{P}_0} \int \phi dP = \alpha$$

2. if ϕ' is over level α ,

$$\sup_{P \in \mathcal{P}_0} \int \phi' dP \leq \alpha$$

then we have that

$$\int \phi dP \geq \int \phi' dP \quad \forall P \in \mathcal{P}$$

In the parametric setting,

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$$

$$\Theta = \Theta_0 \cup \Theta_1, \Theta_0 \cap \Theta_1 = \emptyset$$

$$H_0 : \theta \in \Theta_0 \quad H_1 : \theta \in \Theta_1$$

$$(a) \sup_{\theta \in \Theta_0} \int \phi dP_\theta$$

(b)

$$\int \phi dP_\theta \geq \int \phi' dP_\theta \quad \forall \theta \in \Theta$$

3.3.2 Monotone Likelihood Ratio Assumption

Definition 3.3.2 Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a one parameter family of probability measures, dominated by a σ -definite measure μ . Also let

$$f_\theta = \frac{dP_\theta}{d\mu}$$

The family of densities,

$$\{f_\theta : \theta \in \Theta\}$$

has Monotone Likelihood Ratio (MLR) in $Y(x)$ if for any $\theta_1 < \theta_2$.

$$L(X) = \frac{f_{\theta_2}(x)}{f_{\theta_1}(x)}$$

is monotone (nondecreasing or nonincreasing) function of $Y(x)$ on the set on which $Y(x)$ is defined.

Friday March 17

GET NOTES FROM DHANUSHI

Monday March 20

3.3.3 General Form of UMP-Test for One-Sided Hypothesis

$$H_0 : \theta \leq \theta_0 \quad H_1 : \theta \geq \theta_0$$

$$H_0 : \theta \geq \theta_0 \quad H_1 : \theta \leq \theta_0$$

Theorem 3.3.1 — 3.2 in class. Suppose that $\{f_\theta : \theta \in \Theta \subseteq \mathbb{R}\}$ has MLR (without loss of generality, nondecreasing in $Y(x)$).

1. For $\alpha < 1$ the test of the form

$$\phi(x) = \begin{cases} 1 & Y(x) > k \\ \gamma & Y(x) = k \\ 0 & Y(x) < k \end{cases}$$

combined with, $\int \phi f_{\theta_0} d\mu = \alpha$ is UMP- α for testing (I).

2. If $\alpha < 1$ then the test,

$$\phi'(x) = \begin{cases} 1 & Y(x) > k' \\ \gamma' & Y(x) = k' \\ 0 & Y(x) < k' \end{cases}$$

combined with $\int \phi' f_{\theta_0} d\mu = \alpha$ is UMP- α for testing (II).

The idea is that you want the smoothed $E_{\theta}\phi$ to be extreme. So you want ϕ to be as extreme as possible.

Proof. Step 1. Show test (I) is UMP for testing

$$H_0 : \theta = \theta_0 \quad H_1 : \theta > \theta_0$$

It suffices to show this test is of N-P form for any $\theta > \theta_0$.

So we want to show that test I is of the form,

$$\phi_1(x) = \begin{cases} 1 & f_{\theta}(x) > k_1 f_{\theta_0}(x) \\ 0 & f_{\theta}(x) < k_1 f_{\theta_0}(x) \end{cases}$$

for any $\theta < \theta_0$.

Because $\{f_{\theta} : \theta \in \Theta\}$ has MLR, $\{x : Y(x) \leq k\} \subseteq \{x : L_0(Y(x)) \leq L_0(k)\}$. Thus, $L_0(Y(x)) = L(x)$.

PHOTO

For the same reason, ϕ is 0 on $\{x : Y(x) < k\}$. So ϕ has form ϕ_1 .

Step 2. Want to show that ϕ has size α because ϕ is of form I, nondecreasing in $Y(x)$.

So $\beta_{\phi}(\theta)$ is nondecreasing in θ .

$$\sup_{\theta \leq \theta_0} \beta_{\phi}(\theta) = \beta_{\phi}(\theta_0) = \alpha$$

Step 3. Show that ϕ is UMP- α for

$$H_0 : \theta \leq \theta_0 \quad H_1 : \theta > \theta_0$$

We know that ϕ is UMP among all tests such that $E_{\theta_0}\phi' \in \alpha$ (**) by Step 1.

Now we want ϕ to be UMP among all ϕ' such that

$$\sup_{\theta \leq \theta_0} E_{\theta}\phi'(x) \leq \alpha$$

But (**) is true because statement above implies (**).

The second part of the proof follows by considering

$$\phi = 1 - \psi$$

where ψ is the first test form with size $1 - \alpha$.

■

3.3.4 Properties of UMP- α Test for One Sided Hypotheses

Intuitively, UMP property tells us that the test's (I) power most extreme on higher side.

PHOTO

But it also has to be most extreme on other side.

Corollary 3.1 Suppose that $0 < \beta_{\phi}(\theta_0) < 1$. If ϕ is of the I form, then for any ψ satisfying

$$\beta_\psi(\theta_0) \geq \beta_\phi(\theta_0)$$

we have

$$\beta_\phi(\theta) \leq \beta_\psi(\theta)$$

Proof. From the proof of Theorem 3.2, $1 - \phi$ is UMP for testing form (II) and has size

$$\beta_{1-\phi}(\theta_0) = E_{\theta_0}(1 - \phi) = 1 - \beta_\phi(\theta_0)$$

Thus, for $\theta < \theta_0$,

$$\begin{aligned} \beta_{1-\phi}(\theta) &\geq \beta_{1-\psi}(\theta) \\ 1 - \beta_\phi(\theta) &\geq 1 - \beta_\psi(\theta) \\ \beta_\phi(\theta) &\leq \beta_\psi(\theta) \end{aligned}$$

■

Next property says that if we have test of form (I) then under MLR and indentifiability, then $\beta_\phi(\theta)$ is strictly increasing.

Theorem 3.3.2 — 3.4. Suppose that $\{f_\theta : \theta \in \Theta \subseteq \mathbb{R}\}$ has MLR (nondecreasing in $Y(x)$) and is identifiable. Let us use ϕ of form I such that

$$E_{\theta_0} \phi(x) = \alpha$$

then $\beta_\phi(\theta)$ is strictly increasing on the set $\{\theta : 0 < \beta_\phi(\theta) < 1\}$.

R Review of identifiability. Consider a family of probability measures,

$$\{P_\theta : \theta \in \Theta\}$$

This family is identifiable if whenever $\theta_1 \neq \theta_2$, then $\mathbb{P}_{\theta_1} \neq P_{\theta_2}$.

Suppose GET NOTES

Wednesday March 22

Let $\theta_1 < \theta_2$. As proved in Theorem 3.2, ϕ is of form $(_.-)$:

$$\phi(x) = \begin{cases} 1 & f_{\theta_2} < k' f_{\theta_1} \\ 0 & f_{\theta_2} > k' f_{\theta_1} \end{cases}$$

for some $0 \leq k' \leq \infty$.

So by NPL, ϕ is MP test for

$$H_0 : \theta_1 \quad H_1 : \theta_2$$

and has size $\beta_\phi(\theta_1)$. We know that

$$\beta_\phi(\theta_2) \geq \beta_\phi(\theta_1)$$

So we need to show

$$\beta_\phi(\theta_2) \neq \beta_\phi(\theta_1)$$

Suppose that $\beta_\phi(\theta_2) = \beta_\phi(\theta_1)$ then let $\phi^*(x) \equiv \beta_\phi(\theta_1)$. Then

$$(\phi - \phi^*)(f_{\theta_2} - k'f_{\theta_1}) \geq 0$$

$$\int (\phi - \phi^*)(f_{\theta_2} - k'f_{\theta_1}) d\mu$$

But since we know that the integral is zero, so much the integrand (inside part) be zero. This, combined with $0 < \beta_\phi(\theta_1) < 1$ gives us that

$$\phi \neq \phi^* \quad f_{\theta_2} \neq k'f_{\theta_1}$$

So $\{f_{\theta_2} \neq k'f_{\theta_1}\} \subseteq \{(\phi - \phi^*)(f_{\theta_2} - k'f_{\theta_1}) \neq 0\}$. Note that

$$\mu(f_{\theta_2} \neq k'f_{\theta_1}) = 0$$

GET NOTES

3.4 Uniformly Most Powerful Unbiased Test - 2 Sided Hypothesis

Suppose we want to test

$$H_0 : \theta_1 \leq \theta \leq \theta_2 \quad H_1 : \theta > \theta_2, \theta < \theta_1$$

PHOTO

Bing Li's Notation: -._.-

2nd Case

$$H_0 : \theta \geq \theta_2, \theta \leq \theta_1 \quad H_1 : \theta_1 < \theta < \theta_1$$

Notation: _.-._

We are going to study three types of 2 sided hypotheses

$$\text{I } H_0 : \theta = \theta_0 \quad H_1 : \theta \neq \theta_0$$

$$\text{II } H_0 : \theta_1 \leq \theta \leq \theta_2 \quad H_1 : \theta > \theta_2, \theta < \theta_1$$

$$\text{III } H_0 : \theta \geq \theta_2, \theta \leq \theta_1 \quad H_1 : \theta_1 < \theta < \theta_1$$

The most commonly used is I (then II then III). But the most fundamental proof is III so we'll work backwards to I.

First we show, in the case of I and II that UMP does not exist.

■ **Example 3.1** (No UMP for 2-sided Hyp I)

$$X \sim \text{Bin}(n, \theta)$$

$$0 < \theta < 1$$

$$0 < \alpha < 1$$

$$H_0 : \theta = .5 \quad H_1 : \theta \neq .5$$

First consider, $H_1^+ : \theta > .5$. By Theorem 3.2,

$$\phi_+ = \begin{cases} 1 & x > c_+ \\ \gamma_+ & x = c_+ \\ 0 & x < c_+ \end{cases}$$

with $E_{\theta=.5}(\phi_+) = \alpha$ is UMP- α .

Can similarly, set up ϕ_- for testing $H_1^- : \theta < .5$.

Suppose that ϕ_0 is UMP - α for given hypothesis. Then ϕ_0 is UMP0 α for H_0 vs H_1^+ .

Consequently,

$$E_{\theta}(\phi_0) = E_{\theta}(\phi_+)$$

GET NOTES

■

Friday March 24

3.4.1 Uniformly Most Powerful Unbiased Test (UMPU)

$$H_0 : \mathcal{P}_0 \quad H_1 : \mathcal{P}_1$$

Definition 3.4.1 A test, ϕ , is unbiased for above hypothesis if $\forall P_0 \in \mathcal{P}_0, P_1 \in \mathcal{P}_1$,

$$E_{P_0} \phi(x) \leq E_{P_1} \phi(x)$$

So rejectiong probability is greater if $P \in \mathcal{P}_1$. That is, the probability of rejeciton is higher when you should reject than when you shouldn't.

Definition 3.4.2 A test, ϕ , is called UMPU of size α (UMPU- α) if

1. It has size α
2. For any size- α unbiased ψ , we have

$$\beta_{\phi}(P) \geq \beta_{\psi}(P) \quad \forall P \in \mathcal{P}_0$$

Several simple facts.

Fact 1. UMP- α test is always unibased.

Proof. Suppose ϕ is UMP- α . Let $\psi(x) = \alpha$. (Here we are CWIST-ing, comparing with stupid test.)

Since ϕ is UMP- α , ψ is a tset of size α .

$$E_P \phi(x) \geq \alpha \quad \forall P \in \mathcal{P}_1$$

but for $\alpha = \sup_{P \in \mathcal{P}_0} E_P \phi(x)$,

$$E_{P_1} \phi \geq E_{P_0} \phi \quad \forall P_i \in \mathcal{P}_i, \quad i = \{0, 1\}$$

■

Fact 2. UMPU- α test is always unbaised, CWIST take $\psi(x) = \alpha$. Then we have that

$$E_P \psi(x) \equiv \alpha \quad \forall P$$

So we have that ψ is unbiased of size α .

So

$$E_P \phi(x) \geq E_P \psi(x) = \alpha \quad \forall P \in \mathcal{P}_1$$

But

$$E_P \phi(x) \leq \alpha \quad \forall P \in \mathcal{P}_0$$

So ϕ is unbiased.

Fact 3. A UMP- α test is always a UMPU- α test. (one is larger class/group)

Proposition 3.2 Suppose on $\mathcal{P} = \mathcal{P}_0 \cup \mathcal{P}_1$ is defined a metric, say P , such that

$$P \mapsto E_P \phi(x), \mathcal{P} \rightarrow \mathbb{R}$$

is continuous with respect to P .

If ϕ is size - α unbiased, then

$$E_P \phi(x) \equiv \alpha \quad \forall P \in \bar{\mathcal{P}}_0 \cap \bar{\mathcal{P}}_1$$

Proof. Let $P \in \bar{\mathcal{P}}_0 \cap \bar{\mathcal{P}}_1$. Because $P \in \bar{\mathcal{P}}_0$, there exists $\{P_k\} \subseteq \mathcal{P}_0$ such that

$$\rho(P_k, P) \rightarrow 0, \quad k \rightarrow \infty$$

which implies that

$$E_{P_k} \phi(x) \rightarrow E_P \phi(x)$$

Note that the LHS is less than α since ϕ is unbiased. This implies that

$$E_P \phi(x) \leq \alpha$$

and similarly

$$E_P \phi(x) \geq \alpha$$

so,

$$E_P \phi(x) = \alpha$$

■

3.4.2 Exponential Family

Recall that $\{f_\theta(x) : \theta \in \Theta\}$ is an exponential family with respect to μ if

$$f_\theta(x) = \frac{\exp \eta(\theta) Y(x)}{\int \exp \eta(\theta) Y(x) d\mu(x)}$$

where $\eta = \Theta \rightarrow \mathbb{R}$ is one-to-one.

Another review, analytic functions:

Definition 3.4.3 A function $f : \mathbb{R} \rightarrow \mathbb{R}$ is analytic on an open set G if on any $x \in G$ there exists a neighborhood 0_x such that for all $x' \in 0_x$,

$$f(x') = \sum_{n=0}^{\infty} a_n (x' - x)^n$$

which generalizes a polynomial.

Theorem 3.4.1 If f is analytic, on G and $\{x : f(x) = 0\}$ has a limit point in G then $f(x) = 0$ on G .

Definition 3.4.4 A point x is a limit point of set A , if every neighborhood of x contains a point in A that is not x itself.

Another review,

Lemma 3.3 Suppose $g(x)e^{\theta x}$ is integrable with respect to μ for all $\theta \in \Theta$. Then, in the interior of Θ ,

$$\theta \mapsto \int e^{\theta x} d\mu(x)$$

Futhermore, the derivatives of all order can be taken inside in integral.

Fact: For one parameter exponential family,

$$\dot{\beta}_{\psi}(\theta) = \text{Cov}_{\theta}(Y(X), \psi(X))$$

Proof. GET NOTES

■

Monday March 27

3.4.3 Generalized Neyman Pearson Lemma

Recall island example for intuition.

Lemma 3.4 Suppose f_1, \dots, f_m, f_{m+1} are functions measurable and integrable with respect to μ . Let $C_1, \dots, C_m \in \mathbb{R}$, and

$$\phi^*(x) = \begin{cases} 1 & f_{m+1} > C_1 f_1 + \dots + C_m f_m \\ 0 & f_{m+1} < C_1 f_1 + \dots + C_m f_m \end{cases}$$

Then

1. For any ϕ satisfying

$$\int \phi f_i d\mu = \int \phi^* f_i d\mu \quad i = 1, \dots, m$$

we have

$$\int \phi^* f_{m+1} d\mu \geq \int \phi f_{m+1} d\mu$$

2. If $C_1 \geq 0, C_m \geq 0$ then for any ϕ satisfying

$$\int \phi f_i d\mu \leq \int \phi^* f_i d\mu \quad i = 1, \dots, m$$

we have

$$\int \phi^* f_{m+1} d\mu \geq \int \phi f_{m+1} d\mu$$

Proof. Consider

$$(\phi^* - \phi)(f_{m+1} - C_1 f_1 - \cdots - c_m f_m)$$

Not if second term is greater than zero, then the first must be greater than or equal to zero. Similarly, if second term is less than zero, then so must the first be less than or equal to.

Therefore,

GET NOTES - lots of stupid arrows across the board

■

3.4.4 UMP-Test for Test III

$$H_0 : \theta \leq t_1, \theta \geq t_2 \quad H_1 : t_1 < \theta < t_2$$

We will show that ϕ of the form $(_, \cdot, _)$ is UMP- α for hypothesis III above.

$$\phi(x) = \begin{cases} 1 & t_1 < Y(x) < t_2 \\ \gamma_i & Y(x) = t_i \quad i = 1, 2 \\ 0 & Y(x) < t_1 \text{ or } Y(x) > t_2 \end{cases}$$

Theorem 3.4.2 — 3.6. Let X be a random variable with density

$$f_\theta(x) = c(\theta)e^{\theta Y(x)}$$

Then,

1. for any $\theta_1 < \theta_2$, $0 < \alpha < 1$, there exists $0 < t_1 < t_2 < \infty$, $0 \leq \gamma_1, \gamma_2 \leq 1$ such that this form of $\phi(_, \cdot, _)$ satisfies,

$$E_{\theta_1} \phi(x) = \alpha$$

$$E_{\theta_2} \phi(x) = \alpha$$

2. Also, for any ψ with

$$E_{\theta_1} \psi \leq E_{\theta_1} \phi, E_{\theta_2} \psi \leq E_{\theta_2} \phi$$

we have that for $\theta_1 < \theta < \theta_2$

$$E_\theta \phi(x) \geq E_\theta \psi$$

3. Also, also, let ϕ be $(_, \cdot, _)$ in part 1. Then for any ψ such that

$$E_{\theta_1} \psi(x) = E_{\theta_2} \psi(x) = \alpha$$

we have that for all $\theta < \theta_1$ or $\theta < \theta_2$,

$$E_\theta \phi(x) \leq E_\theta \psi(x)$$

4. Also, also, also, This ϕ is UMP- α for III.

Proof. 1. Proof is similar to existence of $_, \cdot, _$ such that

$$E_{\theta_0}(_, \cdot, _) = \alpha$$

using intermediate value theorem for P. This is tedious, so omitted.

2. Idea: for any $\theta \in (\theta_1, \theta_2)$, $\phi = _$ is of the GNP form.

GET NOTES

3. **Wednesday March 29**

GET NOTES

- 4.



3.4.5 UMPU- α Test for II, III

First, we will address hypothesis II.

Theorem 3.4.3 — 3.7. Suppose X has density

$$\frac{e^{\theta Y(x)}}{\int e^{\theta Y(x)} d\mu(x)} = c(\theta e^{\theta Y(x)})$$

and we know that $\theta_1 < \theta_2$, and $0 < \alpha < 1$. Then

1. there exists test of the form

$$\phi(x) = \begin{cases} 1 & Y(x) < t_1 \text{ or } Y(x) > t_2 \\ \gamma_1 & Y(x) = t_1 \\ \gamma_2 & Y(x) = t_2 \\ 0 & t_1 < Y(x) < t_2 \end{cases}$$

where $-\infty < t_1 < t_2 < \infty$ and $0 \leq \gamma_1, \gamma_2 < 1$ such that

$$E_{\theta_1} \phi = E_{\theta_2} \phi = \alpha$$

2. this test is UMPU- α test for form II

$$H_0 : \theta_1 \leq \theta \leq \theta_2 \quad H_1 : \theta < \theta_1 \text{ or } \theta > \theta_2$$

Proof. 1. Get notes.

- 2.



Theorems 3.6 and 3.7 (the previous two) are logically equivalent. But how come we get UMP for 3.6, but UMPU for 3.7?

For hypothesis II, we need to exclude the test such as

PHOTO

that is why we consider unbiased test.

For hypothesis I, these test are automatically excluded.

Only difference.

Now we may consider hypothesis I.

Theorem 3.4.4 — 3.8. Suppose that X has density,


$$\frac{e^{\theta Y(x)}}{\int e^{\theta Y(x)} d\mu(x)} = c(\theta) e^{\theta Y(x)}$$

Let $\theta_0 \in \Theta^0$ interior of Θ and $0 < \alpha < 1$.

Then,

1. there exists $\phi = \phi(x)$ such that

$$\begin{aligned} E_{\theta_0} \phi(x) &= \alpha \\ E_{\theta_0} [Y(x) \phi(x)] &= \alpha E_{\theta_0} (Y(x)) \end{aligned}$$

 Note that unbiasedness for hyp I implies that (SEE PHOTO)

$$\begin{aligned} \beta_{\phi}(\theta_0) &= \alpha \\ \dot{\beta}_{\phi}(\theta_0) &= 0 \end{aligned}$$

But recall that

$$\dot{\beta}_{\phi}(\theta_0) = \text{Cov}_{\theta_0}(Y(x), \phi(x))$$

which implies that

$$E_{\theta_0}(Y\phi) = \alpha E_{\theta_0} Y$$

2. ϕ is UMPU- α for testing hypothesis I.

Proof. 1. Part 1 is omitted.
2. GET NOTES



Friday March 31

FINISHED PROOF.

■ **Example 3.2** Let X_1, \dots, X_n , iid, $N(\theta, 1)$. We want to test hypotheses I, II, and III.
by elementary calculation (BEC)

$$f_{\theta}(X_1, \dots, X_n) = c(\theta) \exp\left\{\theta \sum_{i=1}^n X_i\right\}$$

So, let $Y(X_1, \dots, X_n) = \sum_{i=1}^n X_i$.

Testing hypothesis III

By Theorem 3.6, we have that the test, ϕ , for hypothesis III is $\phi(x)$ such that

$$E_{\theta_1}(\phi) = E_{\theta_2}(\phi) = \alpha$$

But,

$$Y \sim N(n\theta, n)$$

Note that here, $\phi(x)$ is approximately the same as $\phi(Y)$.
We have that

$$P_{\theta_1}(t_1 < Y < t_2) = \alpha$$

$$P_{\theta_2}(t_1 < Y < t_2) = \alpha$$

Again, by elementary calculations, we know that above are equivalent to

$$\begin{cases} \Phi\left(\frac{t_2 - n\theta_1}{\sqrt{n}}\right) - \Phi\left(\frac{t_1 - n\theta_1}{\sqrt{n}}\right) = \alpha \\ \Phi\left(\frac{t_2 - n\theta_1}{\sqrt{n}}\right) - \Phi\left(\frac{t_1 - n\theta_2}{\sqrt{n}}\right) = \alpha \end{cases}$$

where Φ is the CDF of a Standard Normal distribution.

In principle, we can always solve these two variables equation numerically (such as Newton Raphson). However, in this, the two equations actually reducing to one equation by symmetry.

So,

$$t_1 = n \frac{\theta_1 + \theta_2}{2} - t_0$$

$$t_2 = n \frac{\theta_1 + \theta_2}{2} + t_0$$

And thus, we only have to solve this equation,

$$\Phi\left(\frac{n \frac{\theta_1 + \theta_2}{2} + t_0 - n\theta_1}{\sqrt{n}}\right) - \Phi\left(\frac{n \frac{\theta_1 + \theta_2}{2} - t_0 - n\theta_1}{\sqrt{n}}\right)$$

which reduces to,

$$\Phi\left(\frac{n(\theta_1 + \theta_2) + 2t_0}{2\sqrt{n}}\right) - \Phi\left(\frac{n(\theta_1 + \theta_2) - 2t_0}{2\sqrt{n}}\right)$$

Testing Hypothesis II

$\phi = -_{-} -_{-} -_{-}$

We have that

$$P_{\theta_1}(Y < t_1) + P_{\theta_1}(Y > t_2) = \alpha$$

$$P_{\theta_2}(Y < t_1) + P_{\theta_2}(Y > t_2) = \alpha$$

Again, by elementary calculations, we know that above are equivalent to

$$\begin{cases} \Phi\left(\frac{t_1 - n\theta_1}{\sqrt{n}}\right) + 1 - \Phi\left(\frac{t_2 - n\theta_1}{\sqrt{n}}\right) = \alpha \\ \Phi\left(\frac{t_1 - n\theta_1}{\sqrt{n}}\right) + 1 - \Phi\left(\frac{t_1 - n\theta_2}{\sqrt{n}}\right) = \alpha \end{cases}$$

pHOTOS

So,

$$t_1 = n\left(\frac{\theta_1 + \theta_2}{2}\right) + t_0$$

$$t_2 = n\left(\frac{\theta_1 + \theta_2}{2}\right) - t_0$$

And thus we only need to solve,

$$\Phi\left(\frac{n(\theta_2 - \theta_1) - 2t_0}{2\sqrt{n}}\right) + 1 - \Phi\left(\frac{n(\theta_2 - \theta_1) + 2t_0}{2\sqrt{n}}\right)$$

FINISH FOR HYP III

■

Monday April 3

Need to finish for Hyp III.

Our test is $\phi = \cdot, \cdot$ which has to satisfy

$$\Phi\left(\frac{t_1 - n\theta_0}{\sqrt{(n)}}\right) + 1 - \Phi\left(\frac{t_2 - n\theta_0}{\sqrt{(n)}}\right) = \alpha$$

$$\text{Cov}_{\theta_0} Y, \phi(Y) = 0$$

Now, hope to solve (t_1, t_2) but we may reduce to equation by symmetry

$$\text{Cov}_{\theta_0}(Y, \phi(Y)) = E_{\theta_0}((Y - n\theta_0)\phi(Y)) = 0$$

We want the product of three functions to integrate to zero which would imply that $\phi(Y)$ is symmetric about $n\theta_0$.

So we take

$$t_1 = n\theta_0 - t_0$$

$$t_2 = n\theta_0 + t_0$$

Then we only need to satisfy

$$\Phi\left(\frac{-t_0}{\sqrt{(n)}}\right) + 1 - \Phi\left(\frac{t_0}{\sqrt{(n)}}\right) = \alpha$$

We can solve

$$t_0 = \sqrt{(n)}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

So our UMPU- α test is
PHOTO

4. Testing Hypotheses in the Presense of Nuisance

$\Theta \subseteq \mathbb{R}^p$
 $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$
 $\{\Theta_0, \Theta_1\}$ partition of Θ

$$H_0 : \theta \in \Theta_0 \quad H_1 : \theta \in \Theta_1$$

Want the form of the UMPU-test (in this case we don't have a general UMP form).

More sepecifically interested in testing one parameter (without loss of generality assume θ_1 but we will treat $\theta_{2:p}$ as unknown).

That is

$$H_0 : \theta_1 \in A \quad H_1 : \theta_1 \notin A$$

For example,

$$H_0 : \theta_1 \in a \quad H_1 : \theta_1 \notin a$$

where $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$

We may call θ_1 the parameter of interest, and the others the nuisance parameters.

■ Example 4.1

$$X_1, \dots, X_n \sim N(\mu, \sigma^2)$$

where σ^2 is unknown and we want to test

$$H_0 : \mu = 0 \quad H_1 : \mu \neq 0$$

■

These are the types of problems concerned in this chapter

4.1 Unbiased and Similar Tests

Recall that if $E_\theta \phi$ is in θ then compact on $\bar{\mathcal{P}}_0 \cap \bar{\mathcal{P}}_1$. Specialized to the current setting this result is:

Proposition 4.1: If $\theta \mapsto E_\theta \phi(x)$ is continuous and ϕ is unbiased of size α then

$$E_\theta \phi(x) = \alpha \quad \forall \theta \in \bar{\Theta}_0 \cap \bar{\Theta}_1 = \Theta_B$$

Definition 4.1.1 A test ϕ satisfying $\beta_\phi(\theta) = \alpha \quad \forall \theta, \theta \in \Theta_B$ is called an α similar test.

A test ϕ is called UMP Similar- α test (or UMPS- α test) if

1. ϕ is α similar
2. if ψ is α -similar then

$$E_\theta \phi \geq E_\theta \psi \quad \forall \theta \in \Theta$$

Let \mathcal{U}_α be the collection of all unbiased test of size α . Then we may say that \mathcal{S} represent the collection of all α -similar tests.

By Proposition 4.1, under the condition that $E_\theta \phi(x)$ continuous in θ ,

$$\mathcal{U}_\alpha \subseteq \mathcal{S}_\alpha$$

Theorem 4.1.1 Suppose that $\beta_\phi(\theta)$ is continuous in θ for all $\phi(x)$. If ϕ is UMPS- α and also ϕ has size α then ϕ is UMPU- α .

We need the notion of sufficiency, completeness, and later ancillary. Recall that T is sufficient if $P_\theta(\cdot|T)$ doesn't depend on θ . Also that T is complete if $E_\theta g(T) = 0 \quad \forall \theta$.

Wednesday April 5

Sufficiency, $\{P_\theta, \theta \in \Theta\}, (\Omega, \mathcal{F})$. A statistic T, measurable on \mathcal{F} , is sufficient if for all $A \in \mathcal{F}$, $P_\theta(A|T)$ is the same for all $\theta \in \Theta$.

We say that T is sufficient on $G \subseteq \Theta$ if $P_\theta(A|T)$ is the same for all $\theta \in G$.

This is important when we need to talk about sufficiency for nuisance parameters.

Completeness. A statistic T, measurable on \mathcal{F} , is complete on $G \subseteq \Theta$ if for any $g \in \mathcal{G}$,

$$E_\theta g(T) = 0 \quad \forall \theta \in G$$

which implies that

$$g(T) = 0 \quad a.s. P_\theta \quad \forall \theta \in G$$

T is bddly (bounded?) complete if the above holds for all bounded g measurable T.



In the presence of nuisance parameter, UMP may not exist even for one-sided hypothesis. So we'll always talk about UMPU test here.

Intuition about the theory:

Suppose we want to test

$$H_0 : \theta_1 = a \quad H_1 : \theta_1 \neq a$$

Suppose there exists a statistic T that is sufficient for $\theta_{2:p}$ for each fixed θ . Then the distribution of $P_\theta(X|T)$ is only a function of θ_1 .

$$P_{\theta}(X \in A|T) = P_{\theta_1}(X \in A|T)$$

Now we turned k-parameter problem for (θ) into one-parameter problem (θ_1) . That is, if we condition on statistic sufficient for nuisance parameter, then we get rid of the nuisance parameter.

Futhermore, suppose that the distribution of $X|T$ Exponential Family. Then we may use results to say that the tests above (the dah dah dah tests) are best for hypotheses I, II, III.

But the power that we maximized is the conditional probability, $P_{\theta_1}(X|T)$, while what we want is $P_{\theta}(X)$.

This is why we need completeness.

First, the notion of Neyman - Structure

Definition 4.1.2 A test, $\phi(X)$ has α -Neyman Structure with respect to a statistic, T, if

$$E_{\theta}[\phi(X)|T] = \alpha \text{ (a.s. } P_{\theta}) \quad \forall \theta \in \Theta_B$$

where $\Theta_B = \bar{\Theta}_0 \cap \bar{\Theta}_1$.

We may let $N_{\alpha}(T)$ denote the collection of all ϕ with Neyman Structure with respect to T.

Note that $N_{\alpha}(T) \subseteq \mathcal{S}_{\alpha}$ because,

$$\begin{aligned} \phi \in N_{\alpha}(T) &\Rightarrow E_{\theta}[\phi(X)|T] = \alpha \\ &\Rightarrow E_{\theta}\phi(X) = \alpha \end{aligned}$$

Theorem 4.1.2 — 4.2. If T sufficient and bddly complete for $\theta \in \Theta_B$, then

$$\mathcal{S} \subseteq N_{\alpha}(T)$$

Proof. Suppose $\phi \in \mathcal{S}_{\alpha}$. Then,

$$E_{\theta}\phi(X) = \alpha \quad \forall \theta \in \Theta_B$$

$$E_{\theta}[E_{\theta}(\phi(X)|T) - \alpha] = 0$$

FINISH PROOF ■

We can make a sligtly more general structure,

Corollary 4.1 Suppose that

$$\Theta_B = \bigcup_{r=1}^s \Theta_B^{(r)}$$

and T is SBC on each $\Theta^{(r)}$. Then

$$\mathcal{S}_{\alpha} \subset \mathcal{N}_{\alpha}(T)$$

Proof. ■

Recall that,

$$\mathcal{U}_{\alpha} \subseteq \mathcal{S}_{\alpha} \supseteq \mathcal{N}_{\alpha}(T)$$

Where first requires continuous power condition.

If we have that condition that T SCB either on Θ_B or each $\Theta_B^{(r)}$ then we get,

$$\mathcal{U}_\alpha \subseteq \mathcal{S}_\alpha \subseteq \mathcal{N}_\alpha(T)$$

In either case, under SBC,

$$\mathcal{U}_\alpha \subseteq \mathcal{N}_\alpha(T)$$

4.2 Sufficient and Complete

For components of θ under exponential family, suppose that

$$X \sim \frac{e^{\theta^T t_0(x)}}{\int e^{\theta^T t_0(x)} d\mu_0}$$

and that $X_1, \dots, X_n \sim X$.

Friday April 7

Let

$$X \sim \frac{e^{\theta^T t_0(x)}}{\int e^{\theta^T t_0(x)} d\mu_0(X)}$$

$$(X_1, \dots, X_n) \sim \frac{e^{\theta^T t(x_1, \dots, x_n)}}{\int e^{\theta^T t(x_1, \dots, x_n)} d\mu_0(x_1, \dots, x_n)}$$

Theorem 4.2.1 Suppose we have X_1, \dots, X_n from a exponentail family. Then, $t(X_1, \dots, X_n)$ are complete and sufficient for θ .

$$t(X_1, \dots, X_n) = \sum_{i=1}^n t_0(X_i)$$

Proof. We get sufficiency form Fisher's factorization theorem. Must show completeness.

FINISH PROOF ■

Exponential Family

Marginal distribution is exponential family, so must conditional distribution be.

"Closed" under marginalization and conditioning.

Let (U, V) be a pair of random vectors defined on $(\Omega_U \times \Omega_V, \mathcal{F}_U \times \mathcal{F}_V)$ for any distribution, P_{UV} . Let P_U be the marginal distribution and $P_{V|U}$ be the conditional distribution.

Lemma 4.1 Let Q be a probability measure on $(\Omega_U \times \Omega_V, \mathcal{F}_U \times \mathcal{F}_V)$ and P is another probability measure defined by

$$dP = a(U)b(V)dQ$$

where $a(\cdot) \geq 0, b(\cdot) \geq 0$.

$$\int a(U)b(V)dQ(U, V) = 1$$

Then

$$1. dP_U(u) = a(U)[\int b(V)dQ_{V|U}(v|u)]dQ_U(u)$$

$$2. \quad dP_{V|U}(v|u) = \frac{b(V)dQ_{V|U}(v|u)}{\int b(V')dQ_{V|U}(v'|u)}$$

Proof. 1. FINISH PROOF

2. ■

Now let

$$t(\underline{X}) = \begin{pmatrix} U(\underline{X}) \\ V(\underline{X}) \end{pmatrix}$$

and

$$\theta = \begin{pmatrix} \eta \\ \xi \end{pmatrix}$$

where $\theta^T t(X) = \eta^T U(\underline{X}) + \xi^T V(\underline{X})$

Theorem 4.2.2 Suppose

$$\underline{X} \sim \frac{e^{\theta^T t(\underline{X})}}{\int e^{\theta^T t(\underline{X})} d\mu(\underline{X})}$$

Then

1. For each fixed ξ there exists a measure μ_ξ on $(\Omega_U, \mathcal{F}_U)$ such that

$$\frac{dP_U}{d\mu_\xi} = \frac{e^{\eta^T \eta}}{\int e^{\eta^T \eta} d\mu(U)}$$

2. For each u there exists a measure on $(\Omega_V, \mathcal{F}_V)$ such that

$$\frac{dP_{V|U}(v|u)}{d\mu_u} = \frac{e^{\xi^T V}}{\int e^{\xi^T V} d\mu_u}$$

Monday April 10

Proof. 1.

2. ■

Corollary 4.2 Suppose that

$$(X_1, \dots, X_n) \sim \frac{e^{\theta^T t(X_{1:n})}}{\int e^{\theta^T t(X_{1:n})} d\mu(X_{1:n})}$$

then for each fixed $\theta_1 = a$, $T_{2:p} = t_{2:p}(x_1, \dots, x_n)$ is complete sufficient for

$$\Theta_{(a)} = \{\theta : \theta_1 \leq a\}$$

4.3 UMPU test in the presence of Nuisance Parameter

Lemma 4.2 Suppose that $T = t(X_1, \dots, X_n)$ is sufficient for Θ . Then, for any $\phi(X_1, \dots, X_n)$ there exists a

$$\psi(t(x_1, \dots, x_n)) = \psi_0 t(x_1, \dots, x_n)$$

such that

$$\beta_{\psi_0}(\theta) = \beta_\phi(\theta) \quad \forall \theta$$

Proof. ■

Theorem 4.3.1 — 4.6. Suppose we have X_1, \dots, X_n are iid Exp Family. Then we have

$$\phi_0(z) = (\cdot - |t_{2:p})$$

such that

$$E_n(\phi_0(T)|T_{2:p}) = \alpha$$

is UMPU- α for testing $H_0 : \theta_1 \leq a \quad H_1 : \theta_1 > a$

Theorem 4.3.2 — 4.7. We we have X_1, \dots, X_n is distributed as Exponential Family then for testing:

1. Hypothesis III: $H_0 : \theta_1 \in (a, b)^c \quad H_1 : \theta_1 \in (a, b)$
The test $\phi_0(T) = \cdot - \cdot |t_{2:p}$ with $E_a(\phi_0|t_{2:p}) = E_a(\phi_0|t_{2:p}) = \alpha$ is UMPU- α test.
2. Hypothesis II: $H_0 : \theta_1 \in [a, b] \quad H_1 : \theta_1 \in [a, b]^c$
The test $\phi_0(T) = -\cdot - |t_{2:p}$ with $E_a(\phi|t_{2:p}) = E_a(\phi|t_{2:p}) = \alpha$ is UMPU- α test.
3. Hypothesis I: $H_0 : \theta_1 = a \quad H_1 : \theta_1 \neq a$
The test $\phi_0(T) = -\cdot - |t_{2:p}$ with $E_a(\phi|t_{2:p}) = E_a(\phi|t_{2:p}) = \alpha$ and $E_a(\phi_0(T)T_1|t_{2:p}) = \alpha E_a(T_1|t_{2:p})$ is UMPU- α test.

4.4 Using Basu's Theorem to construct UMPU in presence of nuisance parameters

A technique simplifying UMPU construction so we need the thresholds for the different hypothesis test forms, $t_1, t_2, \gamma_1, \gamma_2$.

These are determined by conditional probability such as

$$E_a(\phi_0|t_{2:p}) = \alpha$$

In special cases, these can be reduced to conditional probabilities which leads to all of the famous tests, student -t, Pearsons test, F-test...

Theorem 4.4.1 — Basu's Theorem. Suppose $X \sim P_\theta$, $T(x)$ is sufficient and bddly complete for θ ; $U(x)$ is ancillary for θ . (This means that the distribution of $U(x)$ doesn't depend on θ .) Then $T(x) \perp\!\!\!\perp U(x)$.