Advanced Statistical Inference

STAT 561 - Advanced Statistical Inference

Dr. Bing Li

PUBLISHED BY PUBLISHER

BOOK-WEBSITE.COM

First printing, March 2013

Contents

I Part One

Part One

1. Basic Ideas in Bayesian Anaylysis

Mathematical Preparation

Monday January 9

1. Product σ-Field

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

   Let $\mathscr{A} = \{AxB : A \in \mathscr{F}_1, B \in \mathscr{F}_2\}$.

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

   Let $E \in \mathscr{F}_1 x \mathscr{F}_2$. 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 $E \in \mathscr{F}_1 x \mathscr{F}_2$ then $E_1(\omega_2) \in \mathscr{F}_1$ for all $\omega_2 \in \Omega_2$. Similarly, $E_2(\omega_1) \in \mathscr{F}_2$ for all $\omega_1 \in \Omega_1$.

   If $f : \Omega_1 x \Omega_2 \to \mathbb{R}$ measurable $\mathscr{F}_1 x \mathscr{F}_2 \setminus \mathscr{R}$. Then for each $\omega_1 \in \Omega_1$,

   $$f(\omega_1, \cdot) \textcircled{m} \mathscr{F}_2 \setminus \mathscr{R} \text{ for each } \omega_2 \in \Omega_2.$$

$$f(\cdot,\omega_2)ⓜ\mathscr{F}_1\setminus\mathscr{R}$$

Now, for each $E\in\mathscr{F}_1x\mathscr{F}_2$ consider

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

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

Proof. Outline.

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

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

  So, $\mu_2(E_2(\omega_1))=\begin{cases}\mu_2(B) & \omega_1\in A\\ \emptyset & \omega_1\notin A\end{cases}=I_A(\omega_1)\mu(B)=f_{1,E}$
  So, $f_{1,E}ⓜ\mathscr{F}_1$.
  Thus $\mathscr{P}\subseteq\mathscr{L}$.
- By $\pi-\lambda$ Theorem, $\mathscr{F}_1x\mathscr{F}_2\subseteq\mathscr{L}$.

Similarly, $f_{2,E}ⓜ\mathscr{F}_2\setminus\mathscr{R}$.
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

$$\pi'=\pi''\text{ on }\mathscr{F}_1x\mathscr{F}_2$$

Note that here, $\mathscr{P}$ equals $\mathscr{A}$ used at begining of notes.
We did not have a measure in $\mathscr{F}_1x\mathscr{F}_2$. Now we have π',π'' both measures on $\mathscr{F}_1x\mathscr{F}_2$, they are the same. We call this measure the product meaure, written as $\mu_1x\mu_2$.
Note that $(\Omega_1x\Omega_2,\mathscr{F}_1x\mathscr{F}_2,\mu_1x\mu_2)$ is called product measure space. ■

3. Tonelli's Theorem

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

   $(\Omega_1x\Omega_2,\mathscr{F}_1x\mathscr{F}_2,\mu_1x\mu_2)$ is the product measure space.

   Suppose we have $f:\Omega_1x\Omega_2\to\mathbb{R}ⓜ\mathscr{F}_1x\mathscr{F}_2\setminus\mathscr{R}$. Where $f\geq 0$ and

$$\int fd(\mu_1x\mu_2)=\int\left[\int(f(\cdot,\omega_2)d\mu_1)\right]d\mu_2$$

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 $P(A|\mathscr{G})$ when $\mathscr{G} \subseteq \mathscr{F}$ as sub-σ-field. We defined this intuitively in elementary probability course (definition above), but we ar enot going to define it generally.

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

$$\nu : \mathscr{G} \to \mathbb{R}, G \mapsto P(AG)$$

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

$$\mu : \mathscr{G} \to \mathbb{R}, G \mapsto P(G)$$

So μ is nothting but P restricted on $\mathscr{G}$.

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

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

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

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

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 $\mathscr{G}$. To distinguish it form P(A|B), where B is a set, we use $P(A||\mathscr{G})$, where $\mathscr{G}$ is a σ-field. By construction,

(a) δ is measurable $\mathscr{G}$
(b) $\int_G \delta dp = P(AG) \quad \forall G \in \mathscr{G}$

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 $P(A||\mathscr{G})$ if and only if (a) and (b) are satisfied. We may define $P(A||\mathscr{G})$ 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) $P(\emptyset||\mathscr{G})=0, P(\Omega||\mathscr{G})=1$ a.s. P
(b) $0\leq P(A||\mathscr{G})\leq 1$ a.s. P
(c) If $A_1, A_2, \ldots$ are disjoint members of $\mathscr{F}$ then $P(\bigcup_n A_n||\mathscr{G})=\sum_n P(A_n||\mathscr{G})$ a.s. P

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

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

Here, we write

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

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

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

(a) f is measurable $\mathscr{G}$
(b) $\int_G fdp=P(AG) \quad \forall G\in\mathscr{P}$

6. Conditional Distribution

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

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

$$f:\mathscr{F}_X x\Omega\rightarrow\mathbb{R}$$

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

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

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

The only condition for this existance is $(\Omega_X,\mathscr{F}_X)$ must be a Borel Space, that is $\mathscr{F}_X$ 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 $\bar{\mathbb{R}}$.

We want to define conditional expectaiton of X given $\mathscr{G}$.

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

$$\nu : \mathscr{G} \to \mathbb{R}, G \mapsto \int_G X dP$$

It can be easily shown that ν is a measure.
Let μ again be $\mathscr{G} \to \mathbb{R}, G \mapsto P(G)$. Then $\nu << \mu$. By RNT, $\delta = \frac{d\nu}{d\mu}$ is well defined. This is defined to be conditional expectation of X given $\mathscr{G}$, written as

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

Suppose $X \ngeq 0$, but integrable P. Recall that $X = X^+ - X^-$. Since $X^+, X^- \geq 0$, then both $E(X^+||\mathscr{G}), E(X^-||\mathscr{G})$ are defined by RNT. We define,

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

Friday January 13
As in the case of $P(A||\mathscr{G})$, the equivalent conditions for $d : \Omega \to \mathbb{R}$ is a version of $E(X||\mathscr{G})$.
(a) δ measurable $\mathscr{G}$
(b) $\int_G \delta dP = \int X dP \quad \forall G \in \mathscr{G}$
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 $A \in \mathscr{F}$, $X = I_A$ then the second condition becomes

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

So, $E(I_A||\mathscr{G}) = P(A||\mathscr{G})$.
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 $E(X||\mathscr{G})$ a.s. P
(b) $a, b \in \mathbb{R}$ then

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

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

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

(d) $|E(X||\mathscr{G})| \leq E(|X|||\mathscr{G})$ a.s. P (in fact this is true for all convex functions).
(e) If $X_n \to X$ a.s. P, $|X_n| \leq Y$, and Y integrable P, then

$$E(X_n||\mathscr{G}) \to E(X||\mathscr{G}) a.s.P$$

Proof. Found in Billingsly. ■

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

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

This is called the Law of Iterative Conditional Expectation.

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

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

Other Properties

(a) X, Y are random elements such that XY integrable P.

(b) If $\mathscr{G} \subseteq \mathscr{F}$ is the sub σ-field, then

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

Conditional expectation is a self-adjoint operation.

Proof. "Wire Theorem"

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

■

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,
$h : \mathscr{F}_X x \Omega_Y \to \mathbb{R}$
This function is called the conditional distribution of X given Y if

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

We say that $\tilde{h} : \mathscr{F}_X x \Omega \to \mathbb{R}$ is the condiitional distribution of X given $\mathscr{G} = Y^{-1}(\mathscr{F}_Y$.
That is,

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

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

(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

$$g : \Omega_X x \Omega_Y$$

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

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

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

$(\Omega, \mathscr{F}, P)$
$(\Omega_X, \mathscr{F}_X, \mu_X)$
$(\Omega_Y, \mathscr{F}_Y, \mu_Y)$
$(\Omega_X x \Omega_Y, \mathscr{F}_X x \mathscr{F}_Y, \mu_X x \mu_Y)$
$(X,Y) : \Omega \to \Omega_X x \Omega_Y ⓜ \mathscr{F} \setminus \mathscr{F}_X x \mathscr{F}_Y$

Let $P_X = PX^{-1}, P_Y = PY^{-1}, P_{XY} = P(XY)^{-1}$.
Assume $P_X << \mu_X, P_Y << \mu_Y, P_{XY} << \mu_X x \mu_Y$.

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

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

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

Let

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

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

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 $A \in \mathscr{F}_X$,

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

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 $G' \in \sigma(y)$

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

But note that

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

So we want to check that

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

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

■

1.1 Frequentist & Bayesian Settings

We have our probability space $(\Omega,\mathscr{F},P)$. We also have some data,

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

$$X:\Omega\to\Omega_X \text{ⓜ} \mathscr{F}/\mathscr{F}_X$$

Here, usually Ω_X is a $\mathbb{R}^m$.

Typically we have

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

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 statical 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, $\mathscr{P}$, where $\mathscr{P} << \mu_x$. Usually we assume that $\mathscr{P}$ is a parametric family, $\mathscr{P} = \{P_\theta : \theta \in \Omega_\theta \subseteq \mathbb{R}^p\}$. We assume that $P_0 \in \mathscr{P}$, that is there exisgts $\theta_0 \in \Omega_\theta$ 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 distribtuion $P_{X|\theta}$. 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 $(\Omega, \mathscr{F}, P)$; σ-finite measurable spaces $(\Omega_X, \mathscr{F}_X, \mu_X)$, $(\Omega_\theta, \mathscr{F}_\theta, \mu_\theta)$. Together,

$$(\Omega_X x \Omega_\theta, \mathscr{F}_X x \mathscr{F}_\theta, \mu_X x \mu_\theta)$$

Also, a random element,

$$(X, \theta) : \Omega \to \Omega_X x \Omega_\theta ⓜ \mathscr{F} \setminus \mathscr{F}_X x \mathscr{F}_\theta$$

$P_X = P \circ X^{-1} \leftarrow$ 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
$P_{X|\theta}(A|\theta) : \mathscr{F}_X x \Omega_\theta \to \mathbb{R}$. Likeliehood distribution
$P_{\theta|X}(G|x) : \mathscr{F}_\theta x \Omega_X \to \mathbb{R}$. Posterior distribution

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

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

$$\pi_{\theta|X} = \frac{f_X \theta}{f_X}$$
$$= \frac{f_{X|\theta} \pi_\theta}{\int_{\Omega_\theta} f_{X|\theta} \pi_\theta d\mu_t heta}$$

1.3 Conditional Independence and Frquentist/Bayesian Sufficiency

Independence

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

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

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

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

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

Also note that

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

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

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

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

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

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

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

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

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

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

Equivalently,

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

Equivalent Condition for Conditional Independence

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

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

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

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

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

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

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

So it suffices to show that

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

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

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

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 $\sigma(\Theta^{-1}(\mathscr{F}_\Theta) \cup \mathscr{G})$.

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

It is trivial to show that $\mathscr{P}$ is a π-system.
MORE IN PHOTO
Meanwhile,

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

Therefore,

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

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

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

So, we want

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

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

Monday January 23

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

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

is conditional probability of

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

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

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

■

1.4 Equivalence of Frequentist & Bayesian Sufficiency

Here we have,

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

Where

$$T:\Omega_X\to\Omega_T \text{ⓜ} \mathscr{F}_X/\mathscr{F}_T$$

is called a statistic.

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

In fewquentist setting, we say that T is **suffienct** 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" fequentist. 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 posterir 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

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

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

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

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 $f_X(x)$?

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

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...

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

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.