S620 - Introduction To statistical theory - Spring 2014 - Information Sheet Final - Enrique Areyan

Exponential Family: iff $f(X_i;\theta) = c(\theta)h(x_i)\exp\left[\sum_{i=1}^{k}\pi_i(\theta)\tau_i(x)\right]$

Lemma: the Joint distribution of $t_1(x),..,t_k(x)$ is of exponential family form with natural parameters $\pi_1(\theta),..,\pi_k(\theta)$. $f(X=(x_1,..,x_n)) = c(\theta)^n\{\prod_{j=1}^{n}h(x_j)\}\exp\left[\sum_{i=1}^{k}\pi_i(\theta)\left(\sum_{j=1}^{n}\tau_i(x_j)\right)\right]$

Transformation Families: Group under function composition Examples: $\mathfrak{X}=\mathbb{R}^p$, $\Theta=\mathbb{R}^p$, $g_\theta(x)=x+\theta$ (location family); $\mathfrak{X}=\mathbb{R}^p$, $\Theta=(0,\infty)$, $g_\theta(x)=\theta x$ (scale family); $\mathfrak{X}=\mathbb{R}$, $\Theta=(-\infty,\infty)\times(0,\infty)$; $g_\theta(x)=\eta+\tau x$; $\theta=(\eta,\tau)$

Sufficiency and completeness: $x\sim f(X;\theta)$, $\theta\in\Theta\subseteq\mathbb{R}^k$; $L_x(\theta)=f(x;\theta)$; $\Lambda_x(\theta_1,\theta_2)=\frac{L_x(\theta_1)}{L_x(\theta_2)}$

Lemma (6.1): Let $t(x)$ be a statistic. TFCAE: (i) $f(x;\theta)=h(x)g(t(x);\theta)$, (ii) if $t(x)=t(y)$, then $\Lambda_x(\theta_1,\theta_2)=\Lambda_y(\theta_1,\theta_2)$ for every $\theta_1,\theta_2\in\Theta$.

Definition (Fisher): the random variable / vector $T(X)$ is sufficient for the parameter θ iff the distribution of $X|T(X)=t$ does not depend on θ.

Factorization theorem: $T(X)$ is sufficient for θ iff $f(X;\theta)=h(x)g(t(x);\theta)$

Note: If $f(x;\theta)$ is of exponential family form, then (6.1) holds with $t(x)=(\tau_1(x),..,\tau_k(x))$

Definition: A sufficient statistic is minimal sufficient iff it is a function of every Suf. statistic

Definition: T is complete iff for any function g the following implication holds

if $E_\theta(g(T))=0\ \forall\theta\in\Theta$ then $P_\theta[g(T)=0]=1\ \forall\theta\in\Theta$

THEOREM: If T is both sufficient and complete then T is minimal sufficient. and $\lambda\in(0,1)$

Def: A function $g:A\to\mathbb{R}$ is convex iff $g(\lambda a_1+(1-\lambda)a_2)\leq\lambda g(a_1)+(1-\lambda)g(a_2)$ $\forall a_1,a_2\in A$ [Ex: x^2 is convex

Jensen's Inequality: X a r.v.; $g:\mathbb{R}\to\mathbb{R}$ convex, then $Eg(X)\geq g(EX)$.

THEOREM 6.3: Estimate a real-valued parameter θ with estimator $d(X)$. Suppose loss $L(\theta,d)$ for each θ. Let $d_1(X)$ be an unbiased estimator for θ and T a sufficient statistic. Then $\chi(T)=E\{d_1(X)|T\}$ estimator is also unbiased. If T is complete (or minimal) $\chi(T)$ is the unique unbiased estimator minimising the risk (MVUE). Rao-Blackwell $\Leftrightarrow L(\theta,d)=(\theta-d)^2$

Likelihood theory: $x\sim f(x;\theta)$; $\theta\in\Theta\subseteq\mathbb{R}^d$; $x=(x_1,..,x_n)$; if $x_1,..,x_n\overset{iid}{\sim}f_1(\cdot;\theta)$, then $f(x;\theta)=\prod_{i=1}^{n}f_1(x_i;\theta)$

Given x: $L(\theta)=L_x(\theta)$ OR $L(\theta;x)=f(x;\theta)$ is the likelihood function and $\ell(\theta)=\log(L(\theta))$ is the log-likelihood function. the likelihood equations are $\nabla_\theta\ell(\theta)=\vec{0}$. Usually, want to find θ^* s.t. (1) $L(\theta^*)\geq L(\theta)\ \forall\theta\in\Theta$ [(2) $\ell(\theta^*)\geq\ell(\theta)\ \forall\theta\in\Theta$: $-\ell(\theta^*)\leq-\ell(\theta^*)\ \forall\theta\in\Theta$ (min)].

Def: Score function $u(\theta)=u(\theta;x)=\nabla_\theta\ell(\theta;x)$, measures how quickly $f(\cdot;\theta)$ varies in θ;

Def: fisher information: $i(\theta)=-E_\theta\left[\frac{d^2}{d\theta^2}\log f(X;\theta)\right]$ hence how readily we can hope to estimate θ.

IID case: $X_1,..,X_n\sim f_1$; $i(\theta)=n\,i_1(\theta)$.

Assume: $\theta\in\Theta\subseteq\mathbb{R}$, $\mathcal{A}=\Theta$; R.V. $Y=W(X)=d(X)$, an estimator of θ. $Z=u(\theta;X)$.

Let $m(\theta)=E_\theta Y$; we know that $EZ=0$. Information inequality: $Var_\theta Y\geq\frac{[m'(\theta)]^2}{i(\theta)}$

Special cases: (1) If $W(X)$ is unbiased for θ then $m(\theta)=\theta$ and $m'(\theta)=1$

(2) If $X_1,..,X_n\sim f_1$ then $i(\theta)=n\,i_1(\theta)$.

Likelihood Ratio tests: Assume that Θ_0 and Θ_1 partition $\Theta\subseteq\mathbb{R}^k$. $H_0:\theta\in\Theta_0$ vs $H_1:\theta\in\Theta_1$

$d=\dim\Theta_0$ [e.g. $H_0=\theta_0$ means that $\Theta_0=\{\theta_0\}$ and $d=\dim\{\theta_0\}=0$]. $p=\dim\Theta$, $m=p-d$.

observe $x=(x_1,..,x_n)$ where $x_1,x_2,..\sim f_1(x_i;\theta)$. $L_0=\sup\{L_x(\theta):\theta\in\Theta_0\}$; $L_1=\sup\{L_x(\theta):\theta\in\Theta\}$

[M.L.E within the respective models]. Obviously, $L_0\leq L_1$; we reject H_0 iff L_1 sufficiently larger than L_0.

Let $T_n=2\log(L_1/L_0)$. theorem: Under suitable regularity conditions $T_n\to\chi^2(m)$.

THEOREM 6.1: A necessary and sufficient condition for a statistic $T(X)$ to be minimal sufficient is $T(x)=T(y)\Leftrightarrow\Lambda_x(\theta_1,\theta_2)=\Lambda_y(\theta_1,\theta_2)\ \forall\theta_1,\theta_2$

EX: Let $\theta = 1,2,3$ index the following three probability distributions for generating $x \in \mathcal{X} = \{1,2,3,4,5\}$:

	X=1	X=2	X=3	X=4	X=5
P_1	0.50	0.35	0.05	0.05	0.05
P_2	0.25	0.20	0.10	0.20	0.25
P_3	0.10	0.20	0.40	0.20	0.10
f_2/f_1	0.5	20/35	2	$\boxed{4}$	$\boxed{5}$
f_3/f_1	0.2	20/35	$\boxed{8}$	$\boxed{4}$	2

a) Let $\alpha = 0.10$. Is there a UMP Test of $H_0: \theta \in \{1\}$ vs $H_1: \theta \in \{2,3\}$? UMP Test that the same test is NP for testing: (i) $H_0: \theta = 1$ vs $H_1: \theta = 2$ and (ii) $H_0: \theta = 1$ vs. $H_1 = \theta = 3$.

(i) NP Test rejects when $x \in \{4,5\}$ (ii) NP Test rejects when $x = \{3,4\}$ not the same test $\Rightarrow$ no UMP TEST

b) Suppose that a prior distribution π assigns prob. 0.50 to $\theta = 1$, 0.25 to $\theta = 2$ and 0.25 to $\theta = 3$. What is the conditional (posterior) distribution of $\theta \mid x = 3$? By def: $\pi(\theta|x) = \pi(\theta)\pi(x|\theta) / \sum_{\theta' \in \Theta} \pi(\theta')\pi(x|\theta')$

Denominator: $\sum_{\theta \in \{1,2,3\}} \pi(X=3|\theta)\pi(\theta) = \pi(x=3|\theta=1)\pi(\theta=1) + \pi(x=3|\theta=2)\pi(\theta=2) + \pi(x=3|\theta)\pi(\theta=3)\} = 0.15$

$\pi(\theta=1|X=3) = (\pi(\theta=1)\pi(X=3|\theta=1)) / 0.15 = 1/6 = \pi(\theta=2|X=3) = (\pi(\theta=2)\pi(X=3|\theta=2)) / 0.15$.

c) The mean of $\pi(\theta|X=3)$ is the best action under square error loss:

$E\begin{bmatrix} 1/6 & \theta=1 \\ 1/6 & \theta=2 \\ 4/6 & \theta=3 \end{bmatrix} = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{4}{6} = \frac{15}{6} = \boxed{\frac{5}{2}}$ Note: this is the same as minimizing bayes rule: $\sum_{\theta \in \{1,2,3\}} L(\theta, d(x))\pi(\theta|X=3)$!

(6.2) Find a minimal sufficient statistic for θ based on an independent sample of size n from the uniform distribution on $(\theta-1, \theta+1)$. Sol: $f(x_i;\theta) = 2^{-1} I(\theta-1 < x_i < \theta+1)$. and $f(x;\theta) = \prod_{i=1}^{n} f(x_i;\theta) = 2^{-n} I(x_{(1)} > \theta-1) I(x_{(n)} < \theta+1)$. $\Rightarrow T(x) = (x_{(1)}, x_{(n)})$ is suff.

(6.3) If $X_1, \dots, X_n \sim Bin(k;\theta)$, then $T(X) = \sum X_i \sim Bin(nk;\theta)$ is a sufficient statistic for θ. It is also complete. A simple unbiased estimator for θ^k, prob. all items are working. By theorem, the MVUE is of θ^k is $I(X_1) = \{1$ if $X_1 = k$, 0 o/w. Then $E[I(X_1)] = \theta^k$. We can compute the distribution of $\chi(T) = E[I(X_1)|T(X)=t] = P_r\{X_1 = k | T(X) = t\}$. $= P_r\{T=t|X_1=k\}P_r\{X_1=k\}/P_r\{T=t\}$

$\chi(T)$: $\chi(T=t) = P_r\{X_1 = k | T = t\} = P_r\{X_1=k, T=t\} / P_r\{T=t\}$ one can solve, $= P_r\{\sum_{i=2}^{n} X_i = t-k\} P_r\{X_1=k\} / P_r\{T=t\}$; Using the fact that $T \sim Bin(nk;\theta)$

$\Rightarrow \boxed{\chi(T) = \binom{kn-k}{t-k} / \binom{kn}{t}}$;

E_1	E_2	...	E_k

Draw $X_1, \dots, X_n \sim \mathcal{X}$. Count the number of occurrences of each E_j

$\Theta = \{\theta \in \mathbb{R}^k : \theta_j \geq 0, \sum_{j=1}^{k}\}$ $\dim \Theta = k-1$; $\Theta_0 = \{\bar{\theta}\}$ $\dim \Theta_0 = 0$: $m = (k-1) - 0 = k-1$.

$L(\theta) = P_\theta$(observe O_1 occurrences of E_1, observe O_2 occurrences of E_2, ..., O_k occurrences of E_k) ; $T_n = 2\log(L_1/L_0)$;

The MLE of θ_j is $\hat{\theta}_j = \frac{O_j}{n}$; $L_1 = L(\hat{\theta})$; $L_0 = L(\bar{\theta})$

$T_n = 2\sum_{j=1}^{k} O_j \log(O_j / \bar{e}_j)$

$Exp(\lambda) = \lambda e^{-\lambda x}$; $Normal(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}}$ $Pois(\mu) = (e^{-\mu}\mu^k)/k!$

$Uniform(0,\theta) = \frac{1}{\theta}$ is not an exponential family [if $f(x;\theta)$ is an exponential family, then the support of $f(x;\theta)$ is the same for every θ]

HYPOTHESIS TESTING: Partition parameter space Θ into Θ_0, Θ_1. Test hypotesis: $H_0: \theta \in \Theta_0$ vs $H_1: \theta \in \Theta_1$. If Θ_i contains a single element $\Theta_i = \{\theta_i\}$, $\theta_i \in \Theta$, then we say that H_i is simple, otherwise H_i is composite

Ex: $X_1, \dots, X_n \sim \text{Normal}(\mu, 1)$. $H_0: \mu = 0$ vs $H_1: \mu \neq 0$. H_0 is simple: $\Theta_0 = \{0\}$.

Ex: $X_1, \dots, X_n \sim \text{Normal}(\mu, \sigma^2)$. $H_0: \mu = 0$ vs $H_1: \mu \neq 0$. H_0 is composite: $\Theta_0 = \{(0, \sigma^2): \sigma^2 \in \mathbb{R}\}$.

In this case σ^2 is a nuisance parameter.

Neyman-Pearson formulation of Hypothesis Testing: (upper bound on prob. of TYPE I error).

Fix $\alpha \in (0,1)$, the significance level. Require that $P_\theta(\text{reject } H_0) \leq \alpha$, $\forall \theta \in \Theta_0$

Test that satisfy this condition are called level-α tests.

A TEST is a function $\phi: \mathcal{X} \to [0,1]$, with the interpretation: if x is observed, then H_0 is rejected with probability $\phi(x)$.

Now, $P_\theta(\text{reject } H_0) = E_\theta \phi(X)$. The size of a test ϕ is: $\sup_{\theta \in \Theta_0} E_\theta \phi(X)$.

A test is level-α iff its size $\leq \alpha$.

the power function of a test ϕ is the function $W: \Theta \to [0,1]$ defined by $W(\theta) = P_\theta(\text{reject } H_0) = E_\theta \phi(X)$. [idea: a good test is one which makes $w(\theta)$ as large as possible on Θ_1, while satisfying $w(\theta) \leq \alpha$, $\forall \theta \in \Theta_0$]

Simple Hypothesis: $H_0: \theta = \theta_0$ vs $H_1: \theta = \theta_1$. Assume densities $f_i(x) = f(x; \theta_i)$.

(Here f_0 is the density when θ_0 is the state of nature). ($f_0 = f_0(x) = f_0(x|\theta_0)$). Define the likelihood ratio $\Lambda(x)$:

$\Lambda(x) = \dfrac{f_1(x)}{f_0(x)}$ (the larger the value of $\Lambda(x)$, the stronger the evidence against H_0).

A Likelihood ratio Test (LRT) of H_0 vs H_1 is a test $\phi_0(x)$ of the form:

$$\phi_0(x) = \begin{cases} 1 & f_1(x) > k f_0(x) \text{, i.e., } \Lambda(x) > k \geq 0 \\ \gamma(x) & f_1(x) = k f_0(x) \text{, where } \gamma: \mathcal{X} \to [0,1] \\ 0 & f_1(x) < k f_0(x) \end{cases}$$

[write example].

[NPL says the test ϕ_0 is the best test of size α]. (Also $\gamma(x) = \gamma_0$)

the critical region of a Test is the set of $x \in \mathcal{X}$ for which $\phi(x) = 1$. (reject H_0), i.e., the critical region is $\phi^{-1}(1)$.

Neyman-Pearson Lemma: (a) Optimality condition (b) Existence condition (c) Uniqueness

Composite Hypotheses: In some situations, use NPL to construct a test that is Uniformly most Powerful (UMP).

Def: A Test ϕ_0 is UMP among level-α tests iff (1) ϕ_0 is a level-α test, i.e., $E_\theta \phi_0(x) \leq \alpha$ $\forall \theta \in \Theta_0$ And (2) if ϕ is another level-α test and $\theta \in \Theta_1$, $E_\theta \phi_0(x) \geq E_\theta \phi(x)$ [ϕ_0 is more powerful than ϕ].

Def: the family of densities $\mathcal{F} = \{f(x; \theta): \theta \in \Theta \subseteq \mathbb{R}\}$ with real scalar parameter θ is of monotone likelihood ratio (MLR) iff there exists a function $t: \mathcal{X} \to \mathbb{R}$ s.t. $\Lambda(x) = f(x; \theta_2)/f(x; \theta_1)$ is a non decreasing function of $t(x)$ for any $\theta_1 \leq \theta_2$.

THEOREM: Suppose that the density of x belongs to a family that has MLR with respect to $t(x)$. For Testing $H_0: \theta \leq \theta_0$ vs $H_1: \theta > \theta_0$, the test

$$\phi_0(x) = \begin{cases} 1 & t(x) > t_0 \\ 0 & t(x) \leq t_0 \end{cases}$$ (need to randomize in discrete case)

is UMP among all tests having the same (or smaller) size.

BAYES FACTORS Posterior odds = Prior odds × Bayes Factor : $\frac{P(\theta_0|x)}{P(\theta_1|x)} = \frac{\pi_0}{\pi_1} \frac{f_0(x)}{f_1(x)}$

where $f_0(x) = f(x|\theta_0)$; $f_1(x) = f(x|\theta_1)$.

BAYES FACTOR $B = \frac{f_0(x)}{f_1(x)}$. Use B to test $H_0: \theta=\theta_0$ vs $H_1: \theta=\theta_1$; reject H_0 iff B is sufficiently smaller than 1. (Test are of the form: reject H_0 iff $B < K$ $\Leftrightarrow$ reject H_0 iff $\frac{f_1(x)}{f_0(x)} > \frac{1}{K} = C$; form of NP test guaranteed by NPL).

Statistical Inference: $X = (X_1, X_2, \ldots, X_n)$. random vector. Joint density $f \in \mathcal{F}$;
$\mathcal{F} = \{f_\theta : \theta \in \Theta \subseteq \mathbb{R}^d\}$; where Θ parameter space.

Types of statistical inference: (1) Point Estimation, (2) Confidence set Estimation, (3) hypothesis testing

Paradigms of statistical inference. (1) Bayesian (prior → posterior dist.) (2) Fisherian (likelihood ratio) (3) Frequentist. (what procedures do well).

Decision theory: 6 elements: (1) Parameter space Θ (states of nature), (2) sample space $\mathfrak{X}$, (3) Family of probability distributions on $\mathfrak{X}$ (4) Action space (5) Loss function (6) a set D of decision rules. (non-randomized decision rule and randomized decision rule).

Risk Function: if state of nature $\theta \in \Theta$ obtains, then the risk associated with non-randomized decision rule d is: $\boxed{R(\theta, d) = E_\theta L(\theta, d(X))} = \begin{cases} \int_x L(\theta, d(x)) f(x;\theta)\,dx \\ \sum_{x \in \mathfrak{X}} L(\theta, d(x)) f(x;\theta) \end{cases}$

Associated with each d is a risk function: $R(\cdot, d) : \Theta \to \mathbb{R}$ as func. (fn. of θ)

KEY NOTION: different decision rules should be compared by comparing their risk function,

Common Loss functions: abs. error $L(\theta, a) = |\theta - a|$; squared error loss $L(\theta, a) = (\theta - a)^2$

Admissibility: let d_1, d_2 be decision rules such that: $R(\theta, d_1) \le R(\theta, d_2)$, $\forall \theta \in \Theta$, with $R(\theta, d_1) < R(\theta, d_2)$ for at least one $\theta \in \Theta$. then, d_1 strictly dominates d_2 $[d_1 > d_2]$

If a decision rule d is strictly dominated by some other decision rule, then d is inadmissible.

If d is not strictly dominated by any other decision rule, then d is admissible.

UMR: Let $\mathcal{D}_0$ be a collection of decision rules. the rule $d^* \in \mathcal{D}_0$ has uniformly minimum risk (UMR) in $\mathcal{D}_0$ iff $\boxed{R(\theta, d^*) \le R(\theta, d) \quad \forall \theta \in \Theta \text{ and } \forall d \in \mathcal{D}_0}$

UMR rules may be impossible to obtain. Two strategies to overcome this difficulty:

I. Impartiality Principles: restrict attention to "reasonable" rules that have constant risk, in which case minimizing the risk function is equivalent to minimizing the single risk value

II. Relax the optimality criterion

I. A. Unbiased rules. A decision rule d is L-unbiased iff:
$E_\theta L(\theta', d(x)) \ge E_\theta L(\theta, d(x)) = R(\theta, d)$, for every $\theta, \theta' \in \Theta$, where θ = true state of nature and θ' = some other state.

theorem: Let Θ be an open subset of $\mathbb{R}^d$. Suppose that $A = \Theta$ [point estimation] and that $L(\theta, a) = \|\theta - a\|^2$ [squared error loss]. then, d is L-unbiased iff $E_\theta d(X) = \theta \ \forall \theta \in \Theta$.

Suppose that $g: \Theta \to \mathbb{R}$, e.g. $g(\theta) = \theta$; Estimate $g(\theta)$ with $L(\theta, a) = [g(\theta) - a]^2$

If $E_\theta d(x) = g(\theta) \ \forall \theta \in \Theta$ then, $R(\theta, d) = E_\theta L(\theta, d(X)) = E_\theta[g(\theta) - d(X)]^2 = Var_\theta d(X)$

UMVU (Uniformly minimum variance among unbiased).

THEOREM: Assume squared error loss and let $Bias_\theta(d) = E_\theta d(x) - \theta$. then

$$risk = variance + (bias)^2$$

II. A. Minimax Principle: $\bar{d}$ is minimax iff $\sup_\theta R(\theta, \bar{d}) \le \sup_\theta R(\theta, d)$, $\forall d \in \mathcal{D}$.

Minimax principle: we should use a minimax decision rule. this is a very conservative rule.

B Bayes Principle: Let π be a probability distribution (set of weights) on Θ.

The Bayes risk of d is $r(\pi, d) = \int_{\theta \in \Theta} R(\theta, d)\, \pi(\theta)\, d\theta$. The Bayes principle, choose d_π, the rule that minimizes the Bayes risk for a given $\pi(\theta)$: $\int_\Theta R(\theta, d_\pi)\, \pi(\theta) \le \int_\Theta R(\theta, d)\pi(\theta)$, $\forall d \in \mathcal{D}$.

The randomized decision rule $d^* = \lambda d_i + (1-\lambda) d_j$ is the rule that takes action $d_i(x)$ with probability λ and action $d_j(x)$ with probability $(1-\lambda)$. the risk function of d^* is: $R(\theta, d^*) = \lambda R(\theta, d_i) + (1-\lambda) R(\theta, d_j)$.

More generally, we allow convex combinations in which case the set of risk functions of all randomized decision rules is the convex hull (smallest convex set containing given points) of the set of risk functions of all nonrandomized decision rules.

THEOREM: Let $\mathcal{D}$ be the class of randomized decision rules. Suppose that d_π is Bayes w.r.t. the prior π. If $\sup_\theta R(\theta, d_\pi) \le r(\pi, d_\pi) = \int_\Theta R(\theta, d_\pi)\pi(d\theta)$, then d_π is minimax.

Admissibility of Bayes Rules.

THEOREM 2.3: Assume that $\Theta = \{\theta_1, .., \theta_t\}$ is finite and $\pi(\cdot)$ is a prob. distr. on Θ. then a Bayes rule w.r.t. π is admissible.

THEOREM 2.4: If a Bayes rule is unique then it is admissible.

THEOREM 2.5: d_π is admissible (continuous case).

BAYESIAN INFERENCE: Treat θ as a r.v. 1) prior distribution on θ 2) inference should be based on posterior distribution.

joint of (θ, x)

$$\theta | x = \frac{\pi(\theta) f(x;\theta)}{\int_\Theta \pi(\theta') f(x;\theta') d\theta'} \text{ [marginal]}$$

$\theta | x \propto$ prior $\times$ likelihood Θ $\theta | x \propto \pi(\theta) f(x;\theta) \Rightarrow$ $[\propto \pi(\theta)\, \pi(x|\theta)]$

$\hookrightarrow$ posterior $\pi(\theta | x)$

To find the Bayes rule, d_π, define $d_\pi(x)$ for each $x \in \mathcal{X}$ in such a way as to minimize the expected posterior loss: $\int_\Theta L(\theta, d(x))\, \pi(\theta | x) d\theta$ as a function of d. distribution $\pi(\theta | x)$.

1) Inferences based on d_π are inferences based on posterior distribution; we need not know
2) Having observed x, we need to know $d(x)$ to take action how to specify the entire decision rule.

$L(\theta, a_i)$	$\theta \in \Theta_0$	$\theta \in \Theta_1$
a_0	0	1 (type II error)
a_1	1 (type I error)	0

CASE 1: Bayesian approach to hypothesis testing

the expected posterior loss is
$$\int_\Theta L(\theta, d(x))\, \pi(\theta|x)\, d\theta = \begin{cases} \int_{\Theta_1} \pi(\theta|x) d\theta & \text{post. prob. of } \Theta_1 \text{ if } d(x) = a_0 \\ \int_{\Theta_0} \pi(\theta|x) d\theta & \text{post. prob. of } \Theta_0 \text{ if } d(x) = a_1 \end{cases}$$

the Bayes rule chooses Θ_i with the larger posterior prob.

CASE 2: Point estimation For squared error loss $L(\theta, a) = (\theta - a)^2$, the expected posterior loss $\int_\Theta [\theta - d(x)]^2 \pi(\theta|x) d\theta$ is minimized by choosing $d(x) = \int_\Theta \theta\, \pi(\theta|x) d\theta$ i.e., the posterior mean [the mean of $\pi(\theta|x)$].

CASE 3 absolute error loss $L(\theta, a) = |\theta - a|$. the Bayes rule is the posterior median of θ

$$\int_{-\infty}^{d} \pi(\theta|x) d\theta = \int_d^{\infty} \pi(\theta|x) d\theta = \frac{1}{2}$$

CASE 4: A BAYESIAN APPROACH TO interval estimation (Primal & Dual approach). [LOOK AT NOTES]

THEOREM: d_π Bayes rule with constant risk $\Rightarrow$ d_π is MINIMAX.

SHRINKAGE and the James-Stein estimator [LOOK at notes]

EMPIRICAL BAYES Approach: estimation of prior parameter values from marginal distributions of data, i.e., use data to estimate hyperparameters.

Bayesian approaches to choosing π (prior): (1) physical models may suggest $\theta \sim \pi$; (2) non-informative priors (3) subjective priors (π expresses beliefs before examining the data) (*) dominant approach (4) do something that is mathematically convenient e.g. a natural conjugate prior.