CH4
Bayesian Analysis

By
Jay Lin

Outline

- The Posterior Distribution
- Bayesian Inference
- Bayesian Decision Theory

1. The Posterior Distribution

Posterior distribution of θ given x = prior info. $(\pi(\theta))$ + sample info. (x)

1.1. Definition and Determination

* posterior dist. = $\pi(\theta| x)$

* joint density fn. = $h(x, \theta) = \pi(\theta) f(x| \theta)$

* marginal density fn. = $m(x) = \int_{\Theta} f(x| \theta)dF^{\pi}(\theta)$

$$\Rightarrow \pi(\theta| x) = \frac{h(x, \theta)}{m(x)}$$

In calculating the posterior dist., it is helpful to use concept of sufficiency.

Lemma 1.

If T is a sufficient statistic for θ with density $g(t| \theta)$. Assume m(t) (the marginal density of t) is greater than zero, and that the factorization theorem holds. Then,

if $T(x) = t \Rightarrow \pi(\theta| x) = \pi(\theta| t) = \dfrac{\pi(\theta)g(t| \theta)}{m(t)}$

Example 1.

Assume $X\sim N(\theta, \sigma^2)$, where θ is unknown but σ^2 is known.

Let $\pi(\theta)$ be a $N(u, \tau^2)$ density, where u and τ^2 are known.

Then $h(x, \theta) = \pi(\theta)f(x| \theta) = (2\pi\sigma\tau)^{-1} \exp\{-\frac{1}{2}[\frac{(\theta-u)^2}{\tau^2}+\frac{(x-\theta)^2}{\sigma^2}]\}$

To find m(x), note that defining $\rho = \dfrac{\tau^2+\sigma^2}{\tau^2\sigma^2}$

then $\frac{1}{2}[\frac{(\theta-u)^2}{\tau^2}+\frac{(x-\theta)^2}{\sigma^2}] = \ldots = \frac{1}{2}\rho[\theta-\frac{1}{\rho}(\frac{u}{\tau^2}+\frac{x}{\sigma^2})]^2+\frac{(u-x)^2}{2(\sigma^2+\tau^2)}$

hence $h(x, \theta) = (2\pi\sigma\tau)^{-1} \exp\{\frac{1}{2}\rho[\theta-\frac{1}{\rho}(\frac{u}{\tau^2}+\frac{x}{\sigma^2})]^2\}\exp\{\frac{(u-x)^2}{2(\sigma^2+\tau^2)}\}$

and $m(x) = \int_{-\infty}^{\infty} h(x, \theta)d\theta = (2\pi\tau)^{-1/2}(\sigma\tau)^{-1} \exp\{-\frac{(u-x)^2}{2(\sigma^2+\tau^2)}\}$

It follows that

$$\pi(\theta|x) = \frac{h(x,\theta)}{m(x)} = \left(\frac{\rho}{2\pi}\right)^{1/2} \exp\left(-\frac{1}{2}\rho\left[\theta - \frac{1}{\rho}\left(\frac{u}{\tau^2} + \frac{x}{\sigma^2}\right)\right]^2\right)$$

$m(x) \sim N(u, (\sigma^2+\tau^2))$ $\quad \pi(\theta|x) \sim N(u(x), \rho^{-1})$

where $u(x) = \frac{1}{\rho}\left(\frac{u}{\tau^2} + \frac{x}{\sigma^2}\right) = x - \frac{\sigma^2}{\sigma^2+\tau^2}(x-u)$

Assume that the test result $X \sim N(\theta, 100)$,

where θ is the true IQ of the child.

For the population, $\theta \sim N(100, 225)$

Using above equaiton, $X \sim N(100, 325)$

$\Rightarrow u(x) = \frac{400+9x}{13}$, $\rho^{-1} = 69.23$

Thus, if a child scores 115, his true IQ $\theta \sim N(110.39, 69.23)$

1.2. Conjugate Families

$m(x)$ and $\pi(\theta | x)$ are not easily calculable.

A large part of the Bayesian literature is devoted to finding prior distributions (Conjugate priors) for which $\pi(\theta| x)$ can be easily calculated.

Definition 1.

Let F denote the class of density functions $f(x| \theta)$. A class P of prior distribution is said to be a conjugate family for F if $\pi(\theta| x)$ is in the class P for all $x \in X$ and $\pi \in P$.

Look back on example 1:

The class of normal priors is a conjugate family, for the class of normal (smaple) densities.

(If X has a normal density and θ has a normal prior, then the posterior density $\pi(\theta| x)$ is also normal).

Example 5

Assume $X_i \sim Poi(\theta)$, $i=1,\dots,n$ $\quad f(x|\theta) = \prod_{i=1}^{n}\left[\frac{\theta^{x_i}e^{-\theta}}{x_i!}\right] = \frac{\theta^{n\bar{x}}e^{-n\theta}}{\prod_{i=1}^{n}[x_i!]}$

Observing that the likelihood fn. for such densities resembles a gamma density. Thus assume $\theta \sim gamma(\alpha, \beta)$

$$h(x,\theta) = f(x|\theta)\pi(\theta) = \frac{e^{-n\theta}\theta^{n\bar{x}}}{\prod_{i=1}^{n}[x_i!]} \frac{\theta^{\alpha-1}e^{-\theta/\beta}I_{(0,\infty)}(\theta)}{\Gamma(\alpha)\beta^{\alpha}} = \frac{e^{-\theta(n+1/\beta)}\theta^{(n\bar{x}+\alpha-1)}I_{(0,\infty)}(\theta)}{\Gamma(\alpha)\beta^{\alpha}\prod_{i=1}^{n}[x_i!]}$$

the factors involoving θ in this expression follow

$gamma(n\bar{x}+\alpha, [n+1/\beta]^{-1}) \Rightarrow \pi(\theta| x)$

$$m(x) = \frac{h(x,\theta)}{\pi(\theta|x)} = \frac{(\Gamma(\alpha)\beta^{\alpha}\prod_{i=1}^{n}[x_i!])^{-1}}{\{\Gamma(\alpha+n\bar{x})[n+1/\beta]^{-(\alpha+n\bar{x})}\}^{-1}}$$

- Conjugate priors:
  – Allowing one to begin with a certain functional form for the prior and end up with a posterior of the same functional form
  – Parameters are updated by the sample info.

2. Bayesian Inference

2.1 Estimation

I. Point estimates

θ estimation: posterior dist. and MLE

Definition 2.

The generalized maximum likelihood estimate of θ is the largest mode, $\hat{\theta}$, of $\pi(\theta| x)$.

$\hat{\theta}$ has the interpretation of being the "most likely" value of θ, given the prior and the sample x.

Example 6

Assume $f(x|\theta) = e^{-(x-\theta)}I_{(\theta,\infty)}(x)$, and $\pi(\theta)=[\pi\cdot(1+\theta^2)]^{-1}$

then $\pi(\theta|x) = \frac{e^{-(x-\theta)}I_{(\theta,\infty)}(x)}{m(x)\cdot\pi\cdot(1+\theta^2)}$ (if $\theta>x$, then $I_{(\theta,\infty)}(x)=0$ and $\pi(\theta| x)=0$)

$$\frac{d}{d\theta}\pi(\theta|x) = \frac{e^{-(x-\theta)}}{m(x)\pi}\left[\frac{e^{\theta}}{1+\theta^2} - \frac{2\theta e^{\theta}}{(1+\theta^2)^2}\right] = \frac{e^{-x}}{m(x)\pi}\frac{e^{\theta}(\theta-1)^2}{(1+\theta^2)^2} > 0$$

$\pi(\theta| x)$ is increasing for $\theta \le x \Rightarrow \hat{\theta} = \max(x)$

In some cases, GMLE must calculated numerically.

II. Estimation Error

Definition 3.

If θ is a real valued parameter with posterior diat. $\pi(\theta|x)$, and δ is the estimate of θ, then the posterior variance of δ is $V_\delta^\pi(x) \equiv E^{\pi(\theta|x)}[(\theta-\delta)^2]$

When δ is the posterior mean $u^\pi(x) \equiv E^{\pi(\theta|x)}[\theta]$

then $V^\pi(x) \equiv V^\pi_{u^\pi}(x)$, and $\sqrt{V^\pi(x)}$ is the posterior standard deviation.

The "standard error" of δ is $\sqrt{V_\delta^\pi(x)}$

$V_\delta^\pi(x) = E^{\pi(\theta|x)}[(\theta-\delta)^2]=E[(\theta-u^\pi(x)+u^\pi(x)-\delta)^2]$

$=...=V^\pi(x)+(u^\pi(x)-\delta)^2$

Look back on example 1.

$V^\pi(x)=\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}$

$x=115 \Rightarrow u^\pi(115)=110.39,\ \sqrt{V^\pi(115)}=8.32$

classical estimate of θ: $\delta=x=115$

$V_\delta^\pi(x)=V^\pi(x)+(u^\pi(x)-x)^2=...=V^\pi(x)+\frac{\sigma^4}{(\sigma^2+\tau^2)^2}(u-x)^2$

$\Rightarrow \sqrt{V_{115}^\pi(115)}=9.49$ (if $(u-x)^2>(\sigma^2+\tau^2)$, then $V_\delta^\pi(x)>\sigma^2$)

$\because \sqrt{V^\pi(115)}=8.32<\sqrt{V_{115}^\pi(115)}=9.49$

We can say that Bayesian estimate is usually better than classical estimate when a reasonable prior is hold.

2.2. Credible Sets

Definition 4.

A 100(1-α)% credible set for θ is a subset C of Θ such that

$$1-\alpha \le P(C|x)=\int_C dF^{\pi(\theta|x)}(\theta)\begin{cases}\int_C \pi(\theta|x)d\theta & \text{(continuous case)}\\ \sum_{\theta\in C}\pi(\theta|x) & \text{(discrete case)}\end{cases}$$

that is the prob. of θ in C.

(the classical confidence procedure is converage prob.)

Definition 5.

The 100(1-α)% HPD (highest posterior density) credible set for θ, is the subset C of Θ of the form

$C=\{\theta\in\Theta: \pi(\theta|x)\ge k(\alpha)\}$, where $k(\alpha)$ is the largest constant such that $P(C|x)\ge 1-\alpha$

From example 1.

The posterior density of θ given x is $N(u(x), \rho^{-1})$.

100(1-α)% HPD => $C=(u(x)+z(\frac{\alpha}{2})\rho^{-1/2}, u(x)-z(\frac{\alpha}{2})\rho^{-1/2})$

x=115 => (94.08, 126.70)

but

a random test score X~N(θ, 100)

=> the classical 95% CI for θ = (95.4, 134.6)

2.3. Hypothesis Testing

Definition 6.

$H_0: \theta\in\Theta_0 \qquad H_1: \theta\in\Theta_1 \qquad \alpha_0=P(\Theta_0|x) \qquad \alpha_1=P(\Theta_1|x)$

π_0 is the prior prob. of Θ_0 $\qquad$ π_1 is the prior prob. of Θ_1

α_0/α_1 is the posterior odds ratio of H_0 to H_1 $\qquad$ π_0/π_1 is the prior odds ratio

$$\Rightarrow B=\frac{\alpha_0/\alpha_1}{\pi_0/\pi_1}=\frac{\alpha_0\pi_1}{\alpha_1\pi_0}$$ (Bayes factor in favor of Θ_0)

interpretation:

$\Theta_0=\{\theta_0\},\ \Theta_1=\{\theta_1\},\ \alpha_0=\frac{\pi_0 f(x|\theta_0)}{\pi_0 f(x|\theta_0)+\pi_1 f(x|\theta_1)},\quad \alpha_1=\frac{\pi_1 f(x|\theta_1)}{\pi_0 f(x|\theta_0)+\pi_1 f(x|\theta_1)}$

$\alpha_0/\alpha_1=\frac{\pi_0 f(x|\theta_0)}{\pi_1 f(x|\theta_1)} \Rightarrow B=\frac{f(x|\theta_0)}{f(x|\theta_1)}$ (likelihood ratio of H_0 to H_1)

in general form:

$$\pi(\theta)=\begin{cases}\pi_0 g_0(\theta) \text{ if } \theta\in\Theta_0\\ \pi_1 g_1(\theta) \text{ if } \theta\in\Theta_1\end{cases}$$

where g_0 and g_1 are densities which describe how the prior mass is spread out over the rwo hypotheses

$$\Rightarrow B=\frac{\int_{\Theta_0} f(x|\theta)dF^{g_0}(\theta)}{\int_{\Theta_1} f(x|\theta)dF^{g_1}(\theta)}$$

if B are insensitive to choices of g_0 and g_1 => "stable" Bayes factor

From example 1:

$H_0: \theta\le 100,\ H_1: \theta>100$

the posterior dist. of θ is N(110.39, 69.23)

then $\alpha_0=P(\theta\le 100|x)=0.106 \quad \alpha_1=P(\theta>100|x)=0.894$

$\Rightarrow \alpha_0/\alpha_1=1/8.44$

the prior dist. of θ is N(100, 225)

$\Rightarrow \pi_0/\pi_1=1$

==> B=1/8.44

I. One-sided Testing

Example 8:

$X \sim N(\theta, \sigma^2)$, θ has the noninformative prior $\pi(\theta)=1$

$\Rightarrow \pi(\theta| x)=N(x, \sigma^2)$

$H_0: \theta \le \theta_0$ $H_1: \theta > \theta_0$

then $\alpha_0=P(\theta \le \theta_0| x)=\Phi((\theta_0-x) / \sigma)$

P-value$=P(X \ge x)=1-\Phi((x-\theta_0) / \sigma)$

(the prob., when $\theta = \theta_0$, of observing an X "more extreme" than the actual data x)

normal dist. is symmetry => α_0 = P-value

* It is possible to justify the use of noninformative prior as an approximation to very vague prior beliefs.
* Any proper prior density which is roughly constant over the interval $(\theta_0-2\sigma, x+2\sigma)$ (assume $x > \theta_0$)
* This may partly explain the preference of many classical practitioner for the use of P-values instead of prob. of Type I and Type II error.

II. Testing a Point Null Hypothesis

More resaonable hypothesis

$\theta \in \Theta_0=(\theta_0-b, \theta_0+b)$

where b>0, and all θ in Θ_0 can be considered "indistinguishable" from θ_0

$\Rightarrow H_0: \theta \in (\theta_0-b, \theta_0+b)$ vs. $H_1: \theta \notin (\theta_0-b, \theta_0+b)$

We need to know when it is suitable to approximate H_0 by $H_0: \theta = \theta_0$.

From Bayesian prospective:

The approximation is reasonable if the posterior prob. of H_0 are nearly equal in the two situaitons.

A very strong condition under when this would be the case is that the obseved likelihood fn. be approximatively constant on (θ_0-b, θ_0+b).

Bayesian test of the point null hypothesis $H_0: \theta = \theta_0$:

We cannot use a continuous prior density.

A reasonable approach is to give θ_0 a postive prob. π_0, while given the $\theta \neq \theta_0$ the density $\pi_1 g_1(\theta)$, where $\pi_1 = 1-\pi_0$ and g_1 is proper. => $H_0: \theta \in (\theta_0-b, \theta_0+b)$

$\because\ m(x)=\int f(x|\theta)dF^{\pi}(\theta)=f(x|\theta_0)\pi_0+(1-\pi_0)m_1(x)$

where $m_1(x)=\int_{\theta\neq\theta_0} f(x|\theta)dF^{g_1}(\theta)$

$$\therefore\ \pi(\theta_0| x) = \frac{f(x|\theta_0)\pi_0}{m(x)} = \frac{f(x|\theta_0)\pi_0}{f(x|\theta_0)\pi_0+(1-\theta_0)m_1(x)} = [1+\frac{(1-\pi_0)}{\pi_0}\frac{m_1(x)}{f(x|\theta_0)}]^{-1}$$

$$\Rightarrow \frac{\alpha_0}{\alpha_1} = \frac{\pi(\theta_0| x)}{1-\pi(\theta_0| x)} = \frac{\pi_0}{\pi_1}\frac{f(x|\theta_0)}{m_1(x)} \Rightarrow B=\frac{f(x|\theta_0)}{m_1(x)}$$

III. Multiple Hypothesis Testing

- From a Bayesian prospective, it is no more difficult to test multiple hypothesis.
- The testing of two hypothesis: just calculate the posterior prob. of each hypothesis.

From Example 1.

The IQ test is to be classified as having below average IQ (less than 90), average IQ (90 to 110), or above average IQ (over 110). Calliing these three regions Θ_1, Θ_2, and Θ_3, and recalling that the posterior is N(110.39, 69.23)

=> $P(\Theta_1| x=115)=0.007$, $P(\Theta_2| x=115)=0.473$, and $P(\Theta_3| x=115)=0.52$

2.4. Predictive Inference

If we try to predict a random variable $Z \sim g(z|\theta)$ based on the observation of $X \sim f(x|\theta)$.

For simplicity, we assume X and Z are indep, otherwise $g(z|\theta)$ would be replaced by $g(z|\theta, x)$.

The idea of Bayesian predictive inference:

1. $\pi(\theta|x)$ is the believed posterior dist. of θ
2. $g(z|\theta)\pi(\theta|x)$ is the joint dist. of z and θ given x
3. then intergrating out over θ will give the believed dist. of z given x.

Definition 7.

The predictive density of Z given x, when the prior for θ is π is definded by $p(z|x)=\int_{\Theta} g(z|\theta)dF^{\pi(\theta|x)}(\theta)$

Example 12.

Consider the linear regression model

$Z = \theta_1+\theta_2 Y+\varepsilon$

where $\varepsilon \sim N(0, \sigma^2)$

A sufficient statistic for $\theta=(\theta_1, \theta_2)$ is the least squares estimation $X=(X_1, X_2)$, where

$$X_2=\sum_{i=1}^{n}(Z_i-\bar{Z})(Y_i-\bar{Y})/\sum_{i=1}^{n}(Y_i-\bar{Y})^2, \quad X_1=\bar{Z}-X_2\bar{Y},$$

Furthermore $X \sim N_2(\theta, \sigma^2\Sigma)$,

where $\Sigma = \frac{1}{SSY}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} y_i^2 & -\bar{y} \\ -\bar{y} & 1 \end{pmatrix}$, $SSY = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$

If the noninformative prior $\pi(\theta)=1$ is used for θ, the posterior dist. $\pi(\theta| x)$, is easily seen to be $N_2(x, \sigma^2\Sigma)$.

We want to predict a future Z corresponding to a given y:

$g(z| \theta)$ is $N(\theta_1+\theta_2 y, \sigma^2)$

=> joint density of (z,θ) given x and y is also normal

=> $p(z|x)$ is the marginal dist. of Z from joint normal posterior

=> $p(z|x) \sim N(x_1+x_2 y, V)$, $V=\sigma^2(1+1/n + (\bar{y}-y)^2/SSY)$

=> $100(1-\alpha)\%$ HPD credible set for Z

$$\left((x_1+x_2 y)+z(\frac{\alpha}{2})\sqrt{V}, (x_1+x_2 y)-z(\frac{\alpha}{2})\sqrt{V}\right)$$

=> noninformative prior Bayesian analyses = classical approach

3. Bayesian Decision Theory

3.1. Posterior Decision Analysis

The conditional Bayes decision principle:

Look at the expected loss of an action for the believed dist. of θ at the time of decision making.

Definition 8.

The posterior expected loss of an action a, when the posterior dist. is $\pi(\theta|x)$, is

$$\rho(\pi(\theta|x), a) = \int_\Theta L(\theta, a) dF^{\pi(\theta|x)}(\theta)$$

A posterior Bayes action $\delta^\pi(x)$ is any action $a \in A$ which minimizes $\rho(\pi(\theta|x), a)$, or equivalently which minimizes $\int_\Theta L(\theta, a) f(x|\theta) dF^\pi(\theta)$

(m(x) need not be calculated)

Another kind of Bayesian anaylsis:

The rule is to minimize Bayes risk, $r(\pi, \delta)$.

Result 1.

A Bayes rule δ^π (minimizing $r(\pi, \delta)$) can be found by choosing, for each x such that $m(x)>0$, an action which minimizes the posterior expected loss.

The rule can be defined arbitrarily when $m(x)=0$.

* Bayes actions need not be unique.
* If $r(\pi, \delta)=\infty$ for all δ, then any decison rule is a Bayes rule.

Result 2.

If δ is a nonrandomized estimator, then

$$r(\pi, \delta) = \int_{\{x:\, m(x)>0\}} \rho(\pi(\theta|x), \delta(x)) dF^m(x)$$

by definition

$$r(\pi, \delta) = \int_\Theta R(\theta, \delta) dF^\pi(\theta) = \int_\Theta \int_X L(\theta, \delta(x)) dF^{X|\theta}(x) dF^\pi(\theta)$$

since $L(\theta, a) \geq -K > -\infty$ and all measures above are finite

Fubini's theorem:

$$r(\pi, \delta) = \begin{cases} \int_X [\int_\Theta L(\theta, \delta(x)) f(x|\theta) dF^\pi(\theta)] dx \\ \sum_{x \in X} [\int_\Theta L(\theta, \delta(x)) f(x|\theta) dF^\pi(\theta)] \end{cases}$$

Definition 9.

If π is an improper prior, but $\delta^\pi(x)$ is an action which minimizes the posterior expected loss of an action for each x with $m(x)>0$, then δ^π is called a generalized Bayes rule.

Bayesian decision theory is by no means limited to standard applications. The methology of choosing a prior and loss and minimizing the posterior expected loss can be applied in almost any situation.

3.2. Estimation

Square – error loss: $L(\theta, a)=(\theta - a)^2$

the posterior expected loss: $\int_\Theta (\theta - a)^2 dF^{\pi(\theta|x)}(\theta)$

=> find out the value of a which minimizing the posterior expected loss.

$0=\frac{d}{da}[\int_\Theta \theta^2 dF^{\pi(\theta|x)}(\theta) - 2a\int_\Theta \theta dF^{\pi(\theta|x)}(\theta) + a^2\int_\Theta dF^{\pi(\theta|x)}(\theta)]$

$= -2E^{\pi(\theta|x)}[\theta] + 2a \quad => a = E^{\pi(\theta|x)}[\theta]$

soloving for a gives the following result:

Result 3.

If $L(\theta,a)=(\theta-a)^2$, the Bayes rule is $\delta^\pi(x) = E^{\pi(\theta|x)}[\theta]$

Result 4.

If $L(\theta,a)=w(\theta)(\theta-a)^2$, the Bayes rule is

$\delta^\pi(x) = \frac{E^{\pi(\theta|x)}[\theta w(\theta)]}{E^{\pi(\theta|x)}[w(\theta)]}$

Resule 5.

If $L(\theta,a)=|\theta-a|$, any median of $\pi(\theta|x)$ is a Bayes estimator of θ

Result 6.

If $L(\theta,a)=\begin{cases} K_0(\theta-a) & \text{if } \theta-a\ge 0 \\ K_1(a-\theta) & \text{if } \theta-a<0 \end{cases}$

any $(K_0/(K_0+K_1))$-fractile of $\pi(\theta|x)$ is a Bayes estimate of θ

From Example 1.

In the estimating the child's IQ, it is deemed to be twice as harmful to underestimate as to overestimate.

=> $K_0 = 2$ and $K_1 = 1$, the 2/3-fractile of a N(0,1) is 0.43,

=> the 2/3-fractile of a N(110.39, 61.23) is

$110.39 + (0.43)(61.23)^{0.5} = 113.97$

3.3. Finite Action Problems and Hypothesis Testing

If $\{a_1,..., a_k\}$ are the available actions and $L(\theta, a_i)$ the corresponding losses, the Bayes action is simply that for which the posterior expected loss $E^{\pi(\theta|x)}[L(\theta,a_i)]$ is the smallest.

$H_0 : \theta\in\Theta_0 \quad H_1 : \theta\in\Theta_1$

If $\theta\in\Theta_0$, then action a_0 is appropriate,while if $\theta\in\Theta_1$, then a_1 is the best.

When the loss is "0-1" ($L(\theta, a_0)=0$, if $\theta\in\Theta_0$ and $L(\theta, a_0)=1$, if $\theta\in\Theta_1$)

=> $E^{\pi(\theta|x)}[L(\theta,a_1)] = \int L(\theta, a_1)dF^{\pi(\theta|x)}(\theta) = \int_{\Theta_0} dF^{\pi(\theta|x)}(\theta) = P(\Theta_0 | x)$,

and $E^{\pi(\theta|x)}[L(\theta,a_0)] = P(\Theta_1 | x)$

For the more realistic "0-K_i" loss:

$L(\theta, a_i)=\begin{cases} 0 & \text{if } \theta\in\Theta_i \\ K_i & \text{if } \theta\in\Theta_j \ (i\ne j) \end{cases}$

the posterior expected loss of $a_0 = K_0P(\Theta_1 | x)$

$a_1 = K_1P(\Theta_0 | x)$

=> if $K_0P(\Theta_1 | x) > K_1P(\Theta_0 | x)$ (a_1 is taken)

=> $\frac{K_0}{K_1} > \frac{P(\Theta_0 | x)}{P(\Theta_1 | x)} = \frac{1-P(\Theta_1 | x)}{P(\Theta_1 | x)}$

=> $C = \{P(\Theta_1 | x) > \frac{K_1}{K_0+K_1}\}$

(rejection region of Bayesian test is the same form as classical likelihood ratio test)

From Example 1.

Bayesian test:

If $x < \theta_0 + \frac{\sigma^2}{\tau^2}(\theta_0 - u) - \sigma^2\rho^{1/2}z(\frac{K_1}{K_1+K_0})$, then reject H_0

classical test:

If $x < \theta_0 + \sigma z(\alpha)$, then then reject H_0

In classical testing, the "critical value" of the rejection region is determined by α, while in the Bayesian test it is determined by the loss and prior information.

Bayesian test can be thought of as providing a rational way of choosing the size of the test. Classical statistics provides no such guildline, with the result being that certain "standard" sizes (0.1, 0.05, 0.01) are ad hoc choices.

Recall the Example 1.

$$\text{If } L(\theta, a_1) = \begin{cases} 0 & \text{if } \theta < 90 \\ \theta\text{-}90 & \text{if } 90 \le \theta \le 110 \\ 2(\theta\text{-}90) & \text{if } \theta > 110 \end{cases} \quad L(\theta, a_2) = \begin{cases} 90-\theta & \text{if } \theta < 90 \\ 0 & \text{if } 90 \le \theta \le 110 \\ \theta\text{-}110 & \text{if } \theta > 110 \end{cases}$$

$$L(\theta, a_3) = \begin{cases} 2(110\text{-}\theta) & \text{if } \theta < 90 \\ 110-\theta & \text{if } 90 \le \theta \le 110 \\ 0 & \text{if } \theta > 110 \end{cases}$$

$$\Rightarrow E^{\pi(\theta|x)}[L(\theta,a_1)] = 34.32,\; E^{\pi(\theta|x)}[L(\theta,a_2)] = 3.55,$$

$$E^{\pi(\theta|x)}[L(\theta,a_3)] = 3.27 \quad \text{thus } a_3 \text{ is the Bayes decision}$$

3.4. With Inference Losses

- Decision theory can be useful for inference problems:
  - First, many common inference measures can be given a formal decision-theoretic representation.
    1. such "0-1" loss in testing leads to standard Bayesian testing measure.
    2. $L(\theta,\ C(x))=1\text{-}I_{C(x)}(\theta)$, where $C(x) \subset \Theta$

- Second, the use of loss fn. to represent the actual success of an inference in communicating info.

$$\alpha(x) = \text{"confidence" with } \theta \text{ in } C(x)$$

$$L_C(\theta,\ \alpha(x))=(I_{C(x)}(\theta)-\alpha(x))^2$$

$$\Rightarrow \alpha^{\pi}(x)=E^{\pi(\theta|x)}[I_{C(x)}(\theta)]=P^{\pi(\theta|x)}(\theta \in C(x))$$

To be continued