

Lecture VII: Bayesian Decision Making – Point Estimation

Charles B. Moss

May 10, 2013

I. The Loss Function and the Utility Function.

- A. Bayesian decision making has been likened to economic decision making under risk and uncertainty. Specifically, economics assumes that individuals make decisions that maximize their expected utility (i.e., the von Neumann-Morgenstern utility formulation).
- B. The assumption is that the decision maker in a Bayes action chooses the decision that minimizes the expected value of his loss function.
- C. **Definition 2.4.1** The elements of a *Bayesian decision problem* are an *action* $a \in A \subseteq R^M$ controlled by the decisionmaker, a *loss function* $L(a, \omega)$ depending on the action and a vector of interest $\omega \in \Omega \subseteq R^q$, and a distribution function P for ω . The objective of the decision maker is to minimize the *Bayes risk function*

$$R(a) = E[L(a, \omega)] = \int_{\Omega} L(a, \omega) p(\omega) d\nu(\omega). \quad (1)$$

- 1. The decision maker needs to choose an action a – say the expected value of a parameter.
- 2. The loss function is the criteria used to choose this action a .
- D. **Definition 2.4.2** If the Bayes risk function $R(a)$ exists for all $a \in A$, then any solution $\hat{a} = \arg \min_{a \in A} R(a)$ of the Bayesian decision problem is a *Bayes action*. The associated *Bayes risk* is $R(\hat{a})$.

E. **Definition 2.4.3** The loss function $L(a, \omega)$ is a *quadratic loss function* if

$$L(a, \omega) = (a - \omega)' Q (a - \omega). \quad (2)$$

where Q is a positive definite matrix.

F. **Theorem 2.4.1 Bayes Action with a Quadratic Loss Function** If the loss function is $L(a, \omega) = (a - \omega)' Q (a - \omega)$ and $A = R^q$, then the Bayes action is $\hat{a} = E[\omega]$ and the Bayes risk is $R(\hat{a}) = \text{tr}[Q \text{var}(\omega)]$.

1. The result is fairly straightforward. Defining the risk function as

$$R(a) = E[L(a, \omega)] = \int_{\Omega} (a - \omega)' Q (a - \omega) p(\omega) d\nu(\omega). \quad (3)$$

Taking the derivative using vector notation yields

$$\nabla_a R(a) = \int_{\Omega} 2Q(a - \omega) p(\omega) d\nu(\omega) = 0. \quad (4)$$

Note we are applying Leibniz's rule for the integral. Since the bounds of the integral do not change, the only thing we are interested in is the derivative of the integrand. Given that a is a constant

$$2Qa = 2Q \int_{\Omega} \omega p(\omega) d\nu(\omega). \quad (5)$$

Multiplying by $1/2Q^{-1}$ yields

$$\hat{a} = \int_{\Omega} \omega p(\omega) d\nu(\omega) = E[\omega]. \quad (6)$$

2. "This result is strong and perhaps surprising in two dimensions: (1) the action is the same no matter what the positive definite matrix Q – a change in Q will affect Bayes risk but will leave the Bayes action unchanged; and (2) it is only the mean of ω that matters for the Bayes action (Geweke)."

3. If we replace ω with its mean (i.e., if the action we are interested in is the mean of the distribution), the loss function becomes $L(a, \bar{\omega}) = (a - \bar{\omega}) Q(a - \bar{\omega})$. In addition, the loss minimizing action becomes $\hat{a} = \bar{\omega}$. This fact is sometimes referred to as the *certainty equivalence principle* Simons, H.A. 1956. Dynamic Programming Under Uncertainty with a Quadratic Criterion Function. *Econometrica* 24(1), 74-81.
4. The quadratic loss function may be inappropriate if the consequences of the loss are not symmetric.
5. **Definition 2.4.4** If $a \in A \subseteq R$ and $\omega \in \Omega \subseteq A$, the loss function $L(a, \omega)$ is a *linear-linear loss function* if

$$L(a, \omega) = (1 - q)(a - \omega) I_{-\infty, a}(\omega) + q(\omega - a) I_{a, \infty}(\omega) \quad (7)$$

where $q \in (0, 1)$.

6. **Theorem 2.4.2 Bayes Action with a Linear-Linear Loss Function** If the loss function $L(a, \omega)$ is Equation 7, the random variable ω is absolutely continuous and $A = R$, then the Bayes action is

$$\hat{a} = \{ a : P(\omega \leq a) = q \} \quad (8)$$

and the Bayes risk is

$$R(\hat{a}) = q(1 - q) \cdot [E(\hat{a} - \omega | \omega \leq \hat{a}) + E(\omega - \hat{a} | \omega > \hat{a})]. \quad (9)$$

"The Bayes action *hata* is the q^{th} quantile of the distribution of ω (Geweke)."

7. **Definition 2.4.5** If the distribution of ω is absolutely continuous, the loss function $L(a, \omega; \epsilon)$ is a *zero-one loss function* if $a \in \Omega$ and

$$L(a, \omega; \epsilon) = 1 - I_{N_\epsilon(a)}(\omega), \quad (10)$$

where $N_\epsilon(a)$ is an open ϵ neighborhood of a .

8. **Theorem 2.4.3 Bayes Action with a Zero-One Loss Function** Suppose that $p(\omega)$ is a continuous function with a unique mode at $\omega = \hat{a}$, and $A = R^q$. Let $\hat{a}(\epsilon)$ be the Bayes action for a zero-one loss function. Then $\lim_{\epsilon \rightarrow 0} \hat{a}(\epsilon) = \hat{a}$.

- G. *Bayesian point estimation* is a Bayes action corresponding to a loss function in which the vector of interest ω is the vector of unobservables θ_A .

$$\hat{\theta}_A \arg \min_{\tilde{\theta}_A} \int_{\Theta_A} L(\tilde{\theta}_A, \theta_A) p(\theta_A | y^0, A) d\nu(\theta_A). \quad (11)$$

- H. The solution of the Bayesian decision problems in the three specific cases just considered carries through directly to the specific case of point estimation of the vector of unobservables θ_A .
1. A quadratic loss function $L(\tilde{\theta}_A, \theta_A)$ leads to the posterior mean $\tilde{\theta}_A = E(\theta_A | y^0, A)$.
 2. The linear-linear loss function for a parameter selects the posterior q^{th} quantile.
 3. The limiting zero-one loss function leads to the mode of $p(\theta_A | y^0, A)$ when applied to the entire vector θ_A .

II. Bayesian Point Estimation Using Simulation

- A. As a simple example consider $x \sim N(\alpha, \sigma_x^2)$. Assume that we want to estimate α . My prior is that $\alpha \sim N(\tilde{\alpha}, \sigma_\alpha^2)$.

1. The kernel of the prior becomes

$$p_\alpha(\alpha | \tilde{\alpha}) \propto \exp[-(\alpha - \tilde{\alpha})^2 / 2] \quad (12)$$

2. The kernel of the sample distribution (the distribution of observables) becomes

$$p_x(x | \alpha, \sigma_x^2) \propto \exp \left[-\frac{\sum_{i=1}^N (x_i - \alpha)^2}{2\sigma_x^2} \right] \quad (13)$$

3. The kernel for the overall likelihood function is then

$$p(x, \alpha | \tilde{\alpha}, \sigma_x^2) \propto \exp \left[-\frac{(\alpha - \tilde{\alpha})^2}{2} - \frac{\sum_{i=1}^N (x_i - \alpha)^2}{2\sigma_x^2} \right] \quad (14)$$

4. Using the quadratic loss function to estimate the α parameter then involves the expectation

$$E(\alpha | x^0) \propto \int_{\alpha} \alpha \exp \left[-\frac{(\alpha - \tilde{\alpha})^2}{2} - \frac{\sum_{i=1}^N (x_i^0 - \alpha)^2}{2\sigma_x^2} \right] d\nu(\alpha). \quad (15)$$

To remove the proportionality, we divide by the probability of observing the sample

$$E(\alpha | x^0) = \frac{\int_{\alpha} \alpha \exp \left[-\frac{(\alpha - \tilde{\alpha})^2}{2} - \frac{\sum_{i=1}^N (x_i^0 - \alpha)^2}{2\sigma_x^2} \right] d\nu(\alpha)}{\int_{\alpha} \exp \left[-\frac{(\alpha - \tilde{\alpha})^2}{2} - \frac{\sum_{i=1}^N (x_i^0 - \alpha)^2}{2\sigma_x^2} \right] d\nu(\alpha)}. \quad (16)$$

B. Estimation using simulation

1. As a starting point, consider writing the likelihood function as

$$p(x, \alpha | \tilde{\alpha}, \sigma_x^2) = p(\alpha | \tilde{\alpha}) \times p(x | \alpha, \sigma_x^2)$$

$$= \exp [-(\alpha - \tilde{\alpha})^2 / 2] \times \exp \left[-\frac{\sum_{i=1}^N (x_i - \alpha)^2}{2\sigma_x^2} \right] \quad (17)$$

2. Using this formulation, we integrate the function by drawing a α from the first kernel and α_i

$$\int_{\alpha} \exp [-(\alpha - \tilde{\alpha})^2 / 2] \times \exp \left[-\frac{\sum_{i=1}^N (x_i^0 - \alpha)^2}{2\sigma_x^2} \right] d\nu(\alpha)$$

$$\approx \sum_{s=1}^S \exp \left(-\frac{\sum_{i=1}^N (x_i^0 - \alpha_s)^2}{2\sigma_x^2} \right) \ni \alpha_s \sim N(\tilde{\alpha}, \sigma_{\alpha}^2)$$

(18)

3. Thus, the Bayesian estimation of the mean using the quadratic loss function can then be written as

$$E(\alpha | x^0) = \frac{\sum_{s=1}^S \alpha_s \exp \left(-\frac{\sum_{i=1}^N (x_i^0 - \alpha_s)^2}{2\sigma_x^2} \right)}{\sum_{s=1}^S \exp \left(-\frac{\sum_{i=1}^N (x_i^0 - \alpha_s)^2}{2\sigma_x^2} \right)} \ni \alpha_s \sim N(\tilde{\alpha}, \sigma_{\alpha}^2) \quad (19)$$

- C. Consider drawing a random variable that is normally distributed with a mean of 5 and a variance of 2.5 – in R

```
rdraw <- rnorm(500,mean=5, sd=sqrt(2.5))
```

1. Next, we write the likelihood function for an particular draw of ndraw random variables as

```
like <- function(a) {  
  ll <- 0  
  for (i in 1:ndraw) {  
    ll <- ll + (x[i]-a)^2 }  
  return (exp(-ll/(2*10))) }  
}
```

2. First, assume that we have a sample of only the first ten of the total population

```
x <- rdraw[1:10]  
ndraw <- 10
```

3. Next, we draw 2000 values of α from our sampling distribution (i.e., $N(2.5, 1.0)$). We keep the value of α , the likelihood of α , and the product of α and the likelihood function.

```
for (s in 1:2000) {  
  a <- rnorm(1,mean=2.5,sd=1.0)  
  if (s==1) res <- cbind(a,like(a),a*like(a)) else  
    res <- rbind(res,cbind(a,like(a),a*like(a)))  
}  
  
bmean <- sum(res[,3])/sum(res[,2])  
cmean <- mean(x)  
rres <- cbind(10,cmean,bmean,bmean/cmean)
```

4. Table 1 presents the classical estimate of the mean and the Bayesian estimate of the mean for various sample sizes.
 - a. First, note that the Bayesian estimate is biased toward the prior (which is lower than the true mean).
 - b. The value of this bias declines as the sample size increases – in the limit it converges to the true mean so the estimator is consistent.

Table 1: Sample Estimates for Classical and Bayesian Formulations

Sample Size	Mean		Relative Mean
	Frequency	Bayesian	
10	4.3066	3.3841	0.7858
20	4.6260	3.8756	0.8378
30	5.1837	4.4622	0.8608
40	5.1263	4.5912	0.8956
50	5.1436	4.6844	0.9107
75	5.0588	4.7939	0.9476
100	5.0360	4.7483	0.9429
150	4.9901	4.8308	0.9681
200	4.9417	4.8107	0.9735
300	4.9983	4.8819	0.9767
400	5.0170	4.9876	0.9941
500	5.0277	5.0135	0.9972