

# Math 362: Mathematical Statistics II

Le Chen

le.chen@emory.edu

Emory University  
Atlanta, GA

Last updated on February 6, 2020

Jan. 30th, 2020

## Chapter 5: Estimation

# Chapter 5. Estimation

§ 5.1 Introduction

§ 5.2 Estimating parameters: MLE and MME

§ 5.3 Interval Estimation

§ 5.4 Properties of Estimators

§ 5.5 Minimum-Variance Estimators: The Cramér-Rao Lower Bound

§ 5.6 Sufficient Estimators

§ 5.7 Consistency

**§ 5.8 Bayesian Estimation**

## § 5.8 Bayesian Estimation

**Rationale:** Let  $W$  be an estimator dependent on a parameter  $\theta$ .

1. Frequentists view  $\theta$  as a parameter whose exact value is to be estimated.
2. Bayesians view  $\theta$  is the value of a random variable  $\Theta$ .

One can incorporate our knowledge on  $\Theta$  — the **prior distribution**  $p_{\Theta}(\theta)$  if  $\Theta$  is discrete and  $f_{\Theta}(\theta)$  if  $\Theta$  is continuous — and use Bayes' formula to update our knowledge on  $\Theta$  upon new observation  $W = w$ :

$$g_{\Theta}(\theta|W = w) = \begin{cases} \frac{p_W(w|\Theta = \theta)f_{\Theta}(\theta)}{\mathbb{P}(W = w)} & \text{if } W \text{ is discrete} \\ \frac{f_W(w|\Theta = \theta)f_{\Theta}(\theta)}{f_W(w)} & \text{if } W \text{ is continuous} \end{cases} \quad (1)$$

where  $g_{\Theta}(\theta|W = w)$  is called **posterior distribution** of  $\Theta$ .

## Four cases for computing posterior distribution

$g_{\Theta}(\theta W = w)$	$W$ discrete	$W$ continuous
$\Theta$ discrete	$\frac{p_W(w \Theta = \theta)p_{\Theta}(\theta)}{\sum_i p_W(w \Theta = \theta_i)p_{\Theta}(\theta_i)}$	$\frac{f_W(w \Theta = \theta)p_{\Theta}(\theta)}{\sum_i f_W(w \Theta = \theta_i)p_{\Theta}(\theta_i)}$
$\Theta$ continuous	$\frac{p_W(w \Theta = \theta)f_{\Theta}(\theta)}{\int_{\mathbb{R}} p_W(w \Theta = \theta')f_{\Theta}(\theta')d\theta'}$	$\frac{f_W(w \Theta = \theta)f_{\Theta}(\theta)}{\int_{\mathbb{R}} f_W(w \Theta = \theta')f_{\Theta}(\theta')d\theta'}$

## Gamma distributions

$$\Gamma(r) := \int_0^{\infty} y^{r-1} e^{-y} dy, \quad r > 0.$$

Two parametrizations for **Gamma distributions**:

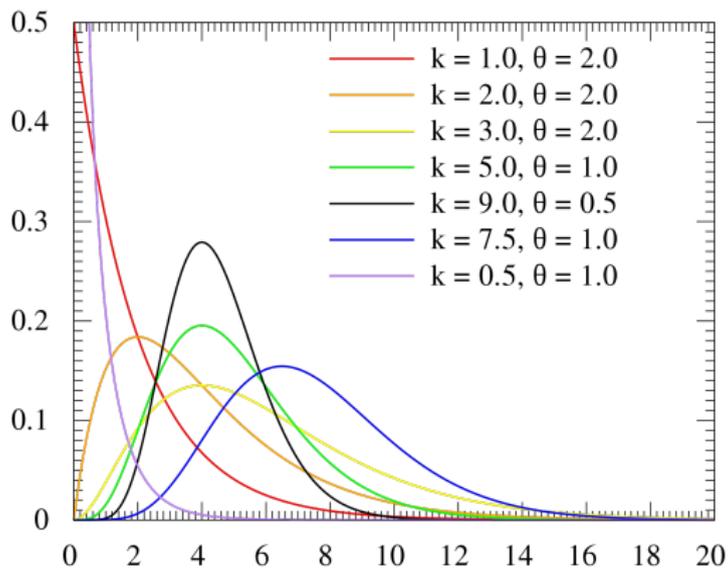
1. With a shape parameter  $r$  and a scale parameter  $\theta$ :

$$f_Y(y; r, \theta) = \frac{y^{r-1} e^{-y/\theta}}{\theta^r \Gamma(r)}, \quad y > 0, r, \theta > 0.$$

2. With a shape parameter  $r$  and a rate parameter  $\lambda = 1/\theta$ ,

$$f_Y(y; r, \lambda) = \frac{\lambda^r y^{r-1} e^{-\lambda y}}{\Gamma(r)}, \quad y > 0, r, \lambda > 0.$$

$$\mathbb{E}[Y] = \frac{r}{\lambda} = r\theta \quad \text{and} \quad \text{Var}(Y) = \frac{r}{\lambda^2} = r\theta^2$$



```

1 # Plot gamma distributions
2 x = seq(0,20,0.01)
3 k= 3 # Shape parameter
4 theta = 0.5 # Scale parameter
5 plot(x,dgamma(x, k, scale = theta),
6       type="l",
7       col="red")
  
```

## Beta distributions

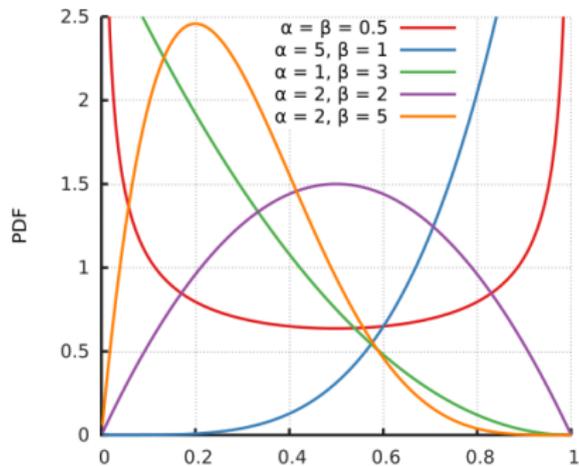
$$B(\alpha, \beta) := \int_0^1 y^{\alpha-1} (1-y)^{\beta-1} dy, \quad \alpha, \beta > 0.$$

$$\begin{aligned} & \vdots \quad \vdots \\ & = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}. \end{aligned} \quad \text{(see Appendix)}$$

Beta distribution

$$f_Y(y; \alpha, \beta) = \frac{y^{\alpha-1} (1-y)^{\beta-1}}{B(\alpha, \beta)}, \quad y \in [0, 1], \alpha, \beta > 0.$$

$$\mathbb{E}[Y] = \frac{\alpha}{\alpha + \beta} \quad \text{and} \quad \text{Var}(Y) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$



```

1 # Plot Beta distributions
2 x = seq(0,1,0.01)
3 a = 13
4 b = 2
5 plot(x, dbeta(x,a,b),
6       type="l",
7       col="red")
  
```

E.g. 1. Let  $X_1, \dots, X_n$  be a random sample from Bernoulli( $\theta$ ):  
 $p_{X_i}(k; \theta) = \theta^k(1 - \theta)^{1-k}$  for  $k = 0, 1$ .

Let  $X = \sum_{i=1}^n X_i$ . Then  $X$  follows binomial( $n, \theta$ ).

Prior distribution:  $\Theta \sim \text{beta}(r, s)$ , i.e.,  $f_{\Theta}(\theta) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \theta^{r-1}(1 - \theta)^{s-1}$  for  $\theta \in [0, 1]$ .

$$\begin{array}{lcl}
 X_1, \dots, X_n \mid \theta & \sim & \text{Bernoulli}(\theta) \\
 \Theta & \sim & \text{Beta}(r, s) \\
 & & r \text{ \& } s \text{ are known}
 \end{array}
 \qquad
 \begin{array}{lcl}
 X = \sum_{i=1}^n X_i \mid \theta & \sim & \text{Binomial}(n, \theta) \\
 \Theta & \sim & \text{Beta}(r, s) \\
 & & r \text{ \& } s \text{ are known}
 \end{array}$$

**Example**  
**5.8.2**

Max, a video game pirate (and Bayesian), is trying to decide how many illegal copies of *Zombie Beach Party* to have on hand for the upcoming holiday season. To get a rough idea of what the demand might be, he talks with  $n$  potential customers and finds that  $X = k$  would buy a copy for a present (or for themselves). The obvious choice for a probability model for  $X$ , of course, would be the binomial pdf. Given  $n$  potential customers, the probability that  $k$  would actually buy one of Max's illegal copies is the familiar

$$p_X(k | \theta) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}, \quad k = 0, 1, \dots, n$$

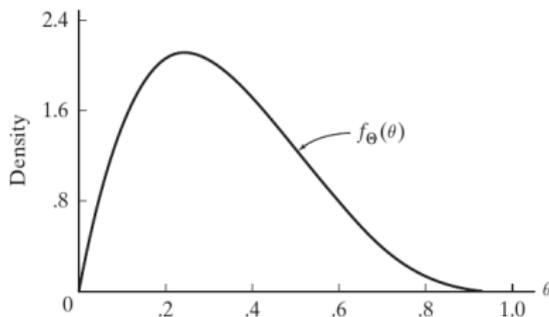
where the maximum likelihood estimate for  $\theta$  is given by  $\theta_e = \frac{k}{n}$ .

It may very well be the case, though, that Max has some additional insight about the value of  $\theta$  on the basis of similar video games that he illegally marketed in previous years. Suppose he suspects, for example, that the percentage of potential customers who will buy *Zombie Beach Party* is likely to be between 3% and 4% and probably will not exceed 7%. A reasonable prior distribution for  $\Theta$ , then, would be a pdf mostly concentrated over the interval 0 to 0.07 with a mean or median in the 0.035 range.

One such probability model whose shape would comply with the restraints that Max is imposing is the *beta pdf*. Written with  $\Theta$  as the random variable, the (two-parameter) beta pdf is given by

$$f_{\Theta}(\theta) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \theta^{r-1} (1-\theta)^{s-1}, \quad 0 \leq \theta \leq 1$$

The beta distribution with  $r = 2$  and  $s = 4$  is pictured in Figure 5.8.1. By choosing different values for  $r$  and  $s$ ,  $f_{\Theta}(\theta)$  can be skewed more sharply to the right or to the left, and the bulk of the distribution can be concentrated close to zero or close to one. The question is, if an appropriate beta pdf is used as a *prior* distribution for  $\Theta$ , and if a random sample of  $k$  potential customers (out of  $n$ ) said they would buy the video game, what would be a reasonable *posterior* distribution for  $\Theta$ ?



**Figure 5.8.1**

$X$  is discrete and  $\Theta$  is continuous.

$$g_{\Theta}(\theta|X = k) = \frac{p_X(k|\Theta = \theta)f_{\Theta}(\theta)}{\int_{\mathbb{R}} p_X(k|\Theta = \theta')f_{\Theta}(\theta')d\theta'}$$

$$\begin{aligned} p_X(k|\Theta = \theta)f_{\Theta}(\theta) &= \binom{n}{k} \theta^k (1 - \theta)^{n-k} \times \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \theta^{r-1} (1 - \theta)^{s-1} \\ &= \binom{n}{k} \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \theta^{k+r-1} (1 - \theta)^{n-k+s-1}, \quad \theta \in [0, 1]. \end{aligned}$$

$$\begin{aligned} p_X(k) &= \int_{\mathbb{R}} p_X(k|\Theta = \theta')f_{\Theta}(\theta')d\theta' \\ &= \binom{n}{k} \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \int_0^1 \theta'^{k+r-1} (1 - \theta')^{n-k+s-1} d\theta' \\ &= \binom{n}{k} \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \times \frac{\Gamma(k+r)\Gamma(n-k+s)}{\Gamma((k+r) + (n-k+s))} \end{aligned}$$

$$\begin{aligned}
 g_{\Theta}(\theta|X = k) &= \frac{\binom{n}{k} \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \times \theta^{k+r-1} (1-\theta)^{n-k+s-1}}{\binom{n}{k} \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \times \frac{\Gamma(k+r)\Gamma(n-k+s)}{\Gamma((k+r) + (n-k+s))}} \\
 &= \frac{\Gamma(n+r+s)}{\Gamma(k+r)\Gamma(n-k+s)} \theta^{k+r-1} (1-\theta)^{n-k+s-1}, \quad \theta \in [0, 1]
 \end{aligned}$$

Conclusion: the posterior  $\sim$  beta distribution( $k+r, n-k+s$ ).

Recall that the prior  $\sim$  beta distribution( $r, s$ ).

It remains to determine the values of  $r$  and  $s$  to incorporate the prior knowledge:

PK 1. Mean is about 0.035.

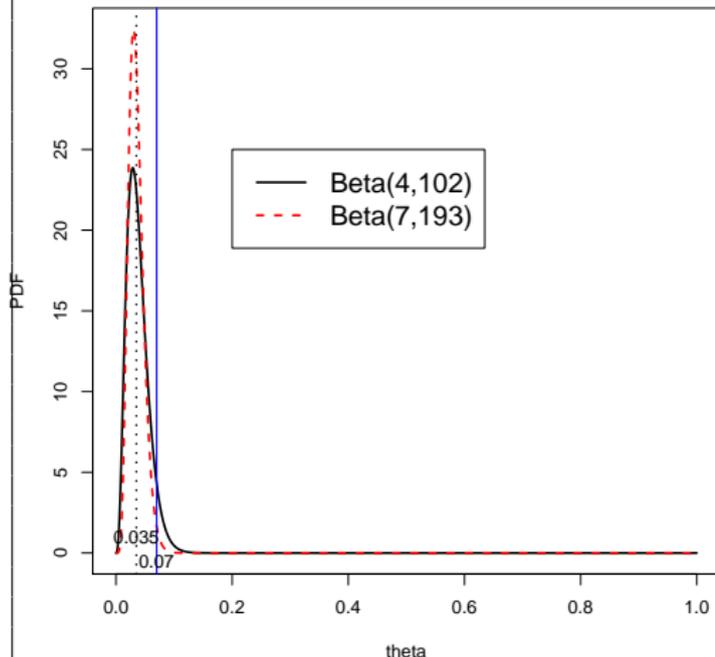
$$\mathbb{E}(\Theta) = 0.035 \implies \frac{r}{r+s} = 0.035 \iff \frac{r}{s} = \frac{7}{193}$$

PK 2. The pdf mostly concentrated over  $[0, 0.07]$ . ... trial ...

```

1 x <- seq(0, 1, length = 1025)
2 plot(x, dbeta(x,4,102),
3      type="l")
4 plot(x, dbeta(x,7,193),
5      type="l")
6 dev.off()
7
8 pdf=cbind(dbeta(x,4,102),dbeta(x,7,193))
9 matplot(x,pdf,
10        type="l",
11        lty = 1:2,
12        xlab = "theta", ylab = "PDF",
13        lwd = 2 # Line width
14        )
15 legend(0.2, 25, # Position of legend
16        c("Beta(4,102)", "Beta(7,193)"),
17        col = 1:2, lty = 1:2,
18        ncol = 1, # Number of columns
19        cex = 1.5, # Fontsize
20        lwd=2 # Line width
21        )
22 abline(v=0.07, col="blue", lty=1,lwd=1.5)
23 text(0.07, -0.5, "0.07")
24 abline(v=0.035, col="gray60", lty=3,lwd=2)
25 text(0.035, 1, "0.035")

```



If we choose  $r = 7$  and  $s = 193$ :

$$g_{\Theta}(\theta|X = k) = \frac{\Gamma(n + 200)}{\Gamma(k + 7)\Gamma(n - k + 193)} \theta^{k+6} (1 - \theta)^{n-k+192}, \quad \theta \in [0, 1]$$

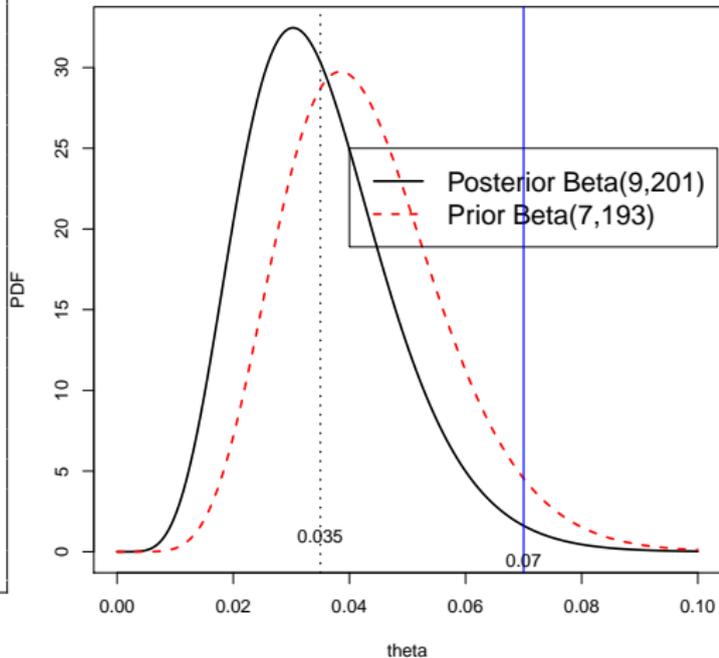
Moreover, if  $n = 10$  and  $k = 2$ ,

$$g_{\Theta}(\theta|X = k) = \frac{\Gamma(210)}{\Gamma(9)\Gamma(201)} \theta^8 (1 - \theta)^{200}, \quad \theta \in [0, 1]$$

```

1 x <- seq(0, 0.1, length = 1025)
2 pdf=cbind(dbeta(x,7,193),dbeta(x,9,201))
3 matplot(x,pdf,
4         type="l",
5         lty = 1:2,
6         xlab = "theta", ylab = "PDF",
7         lwd = 2 # Line width
8     )
9 legend(0.05, 25, # Position of legend
10       c("Posterior Beta(9,201)", "Prior
11         Beta(7,193)"),
12       col = 1:2, lty = 1:2,
13       ncol = 1, # Number of columns
14       cex = 1.5, # Fontsize
15       lwd=2 # Line width
16     )
17 abline(v=0.07,col="blue", lty=1,lwd=1.5)
18 text(0.07, -0.5, "0.07")
19 abline(v=0.035,col="black", lty=3,lwd=2)
20 text(0.035, 1, "0.035")

```



**Definition.** If the posterior distributions  $p(\Theta|X)$  are in the same probability distribution family as the prior probability distribution  $p(\Theta)$ , the prior and posterior are then called **conjugate distributions**, and the prior is called a **conjugate prior** for the likelihood function.

1. Beta distributions are conjugate priors for Bernoulli, binomial, nega. binomial, geometric likelihood.
2. Gamma distributions are conjugate priors for Poisson and exponential likelihood.

E.g. 2. Let  $X_1, \dots, X_n$  be a random sample from  $\text{Poisson}(\theta)$ :  $p_X(k; \theta) = \frac{e^{-\theta} \theta^k}{k!}$  for  $k = 0, 1, \dots$ .

Let  $W = \sum_{i=1}^n X_i$ . Then  $W$  follows  $\text{Poisson}(n\theta)$ .

Prior distribution:  $\Theta \sim \text{Gamma}(s, \mu)$ , i.e.,  $f_\Theta(\theta) = \frac{\mu^s}{\Gamma(s)} \theta^{s-1} e^{-\mu\theta}$  for  $\theta > 0$ .

$$\begin{aligned} X_1, \dots, X_n \mid \theta &\sim \text{Poisson}(\theta) \\ \Theta &\sim \text{Gamma}(s, \mu) \\ &s \ \& \ \mu \ \text{are known} \end{aligned}$$

$$\begin{aligned} W = \sum_{i=1}^n X_i \mid \theta &\sim \text{Poisson}(n\theta) \\ \Theta &\sim \text{Gamma}(s, \mu) \\ &s \ \& \ \mu \ \text{are known} \end{aligned}$$

$$g_{\Theta}(\theta|W = w) = \frac{p_W(w|\Theta = \theta)f_{\Theta}(\theta)}{\int_{\mathbb{R}} p_W(w|\Theta = \theta')f_{\Theta}(\theta')d\theta'}$$

$$\begin{aligned} p_W(w|\Theta = \theta)f_{\Theta}(\theta) &= \frac{e^{-n\theta}(n\theta)^w}{w!} \times \frac{\mu^s}{\Gamma(s)} \theta^{s-1} e^{-\mu\theta} \\ &= \frac{n^w}{w!} \frac{\mu^s}{\Gamma(s)} \times \theta^{w+s-1} e^{-(\mu+n)\theta}, \quad \theta > 0. \end{aligned}$$

$$\begin{aligned} p_W(w) &= \int_{\mathbb{R}} p_W(w|\Theta = \theta')f_{\Theta}(\theta')d\theta' \\ &= \frac{n^w}{w!} \frac{\mu^s}{\Gamma(s)} \int_0^{\infty} \theta'^{w+s-1} e^{-(\mu+n)\theta'} d\theta' \\ &= \frac{n^w}{w!} \frac{\mu^s}{\Gamma(s)} \times \frac{\Gamma(w+s)}{(\mu+n)^{w+s}} \end{aligned}$$

$$\begin{aligned}
 g_{\Theta}(\theta|X = k) &= \frac{\frac{n^w}{w!} \frac{\mu^s}{\Gamma(s)} \times \theta^{w+s-1} e^{-(\mu+n)\theta}}{\frac{n^w}{w!} \frac{\mu^s}{\Gamma(s)} \times \frac{\Gamma(w+s)}{(\mu+n)^{w+s}}} \\
 &= \frac{(\mu+n)^{w+s}}{\Gamma(w+s)} \theta^{w+s-1} e^{-(\mu+n)\theta}, \quad \theta > 0.
 \end{aligned}$$

Conclusion: the posterior of  $\Theta \sim \text{gamma distribution}(w + s, n + \mu)$ .

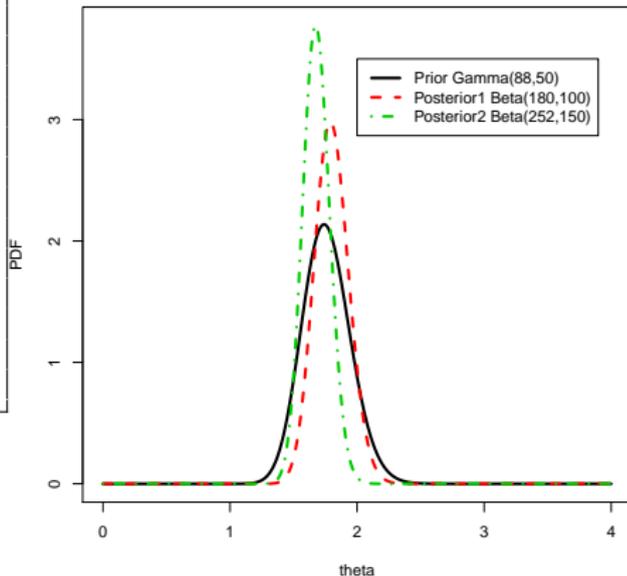
Recall that the prior of  $\Theta \sim \text{gamma distribution}(s, \mu)$ .

## Case Study 5.8.1

```
1 x <- seq(0, 4, length = 1025)
2 pdf=cbind(dgamma(x, shape=88, rate=50),
3           dgamma(x, shape=88+92, 100),
4           dgamma(x, 88+92+72, 150))
5 matplot(x,pdf,
6         type="l",
7         lty = 1:3,
8         xlab = "theta", ylab = "PDF",
9         lwd = 2 # Line width
10 )
11 legend(2, 3.5, # Position of legend
12        c("Prior Gamma(88,50)",
13          "Posterior1 Beta(180,100)",
14          "Posterior2 Beta(252,150)"),
15        col = 1:3, lty = 1:3,
16        ncol = 1, # Number of columns
17        cex = 1.5, # Fontsize
18        lwd=2 # Line width
19 )
```

**Table 5.8.1**

Years	Number of Hurricanes
1851–1900	88
1901–1950	92
1951–2000	72



## Bayesian Point Estimation

**Question.** Can one calculate an appropriate *point estimate*  $\theta_e$  given the posterior  $g_{\Theta}(\theta|W = w)$ ?

**Definitions.** Let  $\theta_e$  be an estimate for  $\theta$  based on a statistic  $W$ . The **loss function** associated with  $\theta_e$  is denoted  $L(\theta_e, \theta)$ , where  $L(\theta_e, \theta) \geq 0$  and  $L(\theta, \theta) = 0$ .

Let  $g_{\Theta}(\theta|W = w)$  be the posterior distribution of the random variable  $\Theta$ . Then the **risk** associated with  $\hat{\theta}$  is the expected value of the loss function with respect to the posterior distribution of  $\Theta$ :

$$\text{risk} = \begin{cases} \int_{\mathbb{R}} L(\hat{\theta}, \theta) g_{\Theta}(\theta|W = w) d\theta & \text{if } \Theta \text{ is continuous} \\ \sum_i L(\hat{\theta}, \theta_i) g_{\Theta}(\theta_i|W = w) & \text{if } \Theta \text{ is discrete} \end{cases}$$

**Theorem.** Let  $g_{\Theta}(\theta|W = w)$  be the posterior distribution of the random variable  $\Theta$ .

- a. If  $L(\theta_e, \theta) = |\theta_e - \theta|$ , then the Bayes point estimate for  $\theta$  is the median of  $g_{\Theta}(\theta|W = w)$ .
- b. If  $L(\theta_e, \theta) = (\theta_e - \theta)^2$ , then the Bayes point estimate for  $\theta$  is the mean of  $g_{\Theta}(\theta|W = w)$ .

Proof. ....



Remarks

- ▶ Mean usually has a closed formula.
- ▶ Median usually does not have a closed form formula.

E.g. 1'.  $X_1, \dots, X_n | \theta \sim \text{Bernoulli}(\theta)$        $X = \sum_{i=1}^n X_i | \theta \sim \text{Binomial}(n, \theta)$

$\Theta \sim \text{Beta}(r, s)$        $\Theta \sim \text{Beta}(r, s)$

$r$  &  $s$  are known       $r$  &  $s$  are known

Prior  $\text{Beta}(r, s) \rightarrow$  posterior  $\text{Beta}(k + r, n - k + s)$   
 upon observing  $X = k$  for a random sample of size  $n$ .

Consider the  $L^2$  loss function.

$$\begin{aligned} \theta_e &= \text{mean of } \text{Beta}(k + r, n - k + s) \\ &= \frac{k + r}{n + r + s} \\ &= \frac{n}{n + r + s} \times \underbrace{\left(\frac{k}{n}\right)}_{\text{MLE}} + \frac{r + s}{n + r + s} \times \underbrace{\left(\frac{r}{r + s}\right)}_{\text{Mean of Prior}} \end{aligned}$$

E.g. 2'.  $X_1, \dots, X_n \mid \theta \sim \text{Poisson}(\theta)$   $W = \sum_{i=1}^n X_i \mid \theta \sim \text{Poisson}(n\theta)$

$\Theta \sim \text{Gamma}(s, \mu)$   $\Theta \sim \text{Gamma}(s, \mu)$

$s$  &  $\mu$  are known  $s$  &  $\mu$  are known

Prior  $\text{Gamma}(s, \mu) \rightarrow$  Posterior  $\text{Gamma}(w + s, \mu + n)$   
 upon observing  $W = w$  for a random sample of size  $n$ .

Consider the  $L^2$  loss function.

$$\begin{aligned} \theta_e &= \text{mean of } \text{Gamma}(w + s, \mu + n) \\ &= \frac{w + s}{\mu + n} \\ &= \frac{n}{\mu + n} \times \underbrace{\left(\frac{w}{n}\right)}_{\text{MLE}} + \frac{\mu}{\mu + n} \times \underbrace{\left(\frac{s}{\mu}\right)}_{\text{Mean of Prior}} \end{aligned}$$

## Appendix: Beta integral

**Lemma.**  $B(\alpha, \beta) := \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

**Proof.** Notice that

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad \text{and} \quad \Gamma(\beta) = \int_0^{\infty} y^{\beta-1} e^{-y} dy.$$

Hence,

$$\Gamma(\alpha)\Gamma(\beta) = \int_0^{\infty} \int_0^{\infty} x^{\alpha-1} y^{\beta-1} e^{-(x+y)} dx dy.$$

The key in the proof is the following change of variables:

$$\begin{cases} x = r^2 \cos^2(\theta) \\ y = r^2 \sin^2(\theta) \end{cases}$$

$$\implies \frac{\partial(x, y)}{\partial(r, \theta)} = \begin{pmatrix} 2r \cos^2(\theta) & 2r \sin^2(\theta) \\ -2r^2 \cos(\theta) \sin(\theta) & 2r^2 \cos(\theta) \sin(\theta) \end{pmatrix}$$

$$\implies \left| \det \left( \frac{\partial(x, y)}{\partial(r, \theta)} \right) \right| = 4r^3 \sin(\theta) \cos(\theta).$$

Therefore,

$$\begin{aligned}\Gamma(\alpha)\Gamma(\beta) &= \int_0^{\frac{\pi}{2}} d\theta \int_0^\infty dr r^{2(\alpha+\beta)-4} e^{-r^2} \cos^{2\alpha-2}(\theta) \sin^{2\beta-2}(\theta) \times \underbrace{4r^3 \sin(\theta) \cos(\theta)}_{\text{Jacobian}} \\ &= 4 \left( \int_0^{\frac{\pi}{2}} \cos^{2\alpha-1}(\theta) \sin^{2\beta-1}(\theta) d\theta \right) \left( \int_0^\infty r^{2(\alpha+\beta)-1} e^{-r^2} dr \right).\end{aligned}$$

Now let us compute the following two integrals separately:

$$I_1 := \int_0^{\frac{\pi}{2}} \cos^{2\alpha-1}(\theta) \sin^{2\beta-1}(\theta) d\theta$$

$$I_2 := \int_0^\infty r^{2(\alpha+\beta)-1} e^{-r^2} dr$$

For  $I_2$ , by change of variable  $r^2 = u$  (so that  $2rdr = du$ ),

$$\begin{aligned} I_2 &= \int_0^\infty r^{2(\alpha+\beta)-1} e^{-r^2} dr \\ &= \frac{1}{2} \int_0^\infty r^{2(\alpha+\beta)-2} e^{-r^2} \underbrace{2rdr}_{=du} \\ &= \frac{1}{2} \int_0^\infty u^{\alpha+\beta-1} e^{-u} du \\ &= \frac{1}{2} \Gamma(\alpha + \beta). \end{aligned}$$

For  $I_1$ , by the change of variables  $\sqrt{x} = \cos(\theta)$  (so that  $-\sin(\theta)d\theta = \frac{1}{2\sqrt{x}}dx$ ),

$$\begin{aligned} I_1 &= \int_0^{\frac{\pi}{2}} \cos^{2\alpha-1}(\theta) \sin^{2\beta-1}(\theta) d\theta \\ &= \int_0^{\frac{\pi}{2}} \cos^{2\alpha-1}(\theta) \sin^{2\beta-2}(\theta) \times \underbrace{\sin(\theta) d\theta}_{=-\frac{1}{2\sqrt{x}} dx} \\ &= \int_1^0 x^{\alpha-\frac{1}{2}} (1-x)^{\beta-1} \frac{-1}{2\sqrt{x}} dx \\ &= \frac{1}{2} \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx \\ &= \frac{1}{2} B(\alpha, \beta) \end{aligned}$$

Therefore,

$$\begin{aligned}\Gamma(\alpha)\Gamma(\beta) &= 4I_1 \times I_2 \\ &= 4 \times \frac{1}{2}\Gamma(\alpha + \beta) \times \frac{1}{2}B(\alpha, \beta)\end{aligned}$$

i.e.,

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

□