Lecture 9

9.1 Prior and posterior distributions.

(Textbook, Sections 6.1 and 6.2)

Assume that the sample $X_1, \ldots, X_n$ is i.i.d. with distribution $\mathbb{P}_{\theta_0}$ that comes from the family $\{\mathbb{P}_\theta : \theta \in \Theta\}$ and we would like to estimate unknown θ_0. So far we have discussed two methods - method of moments and maximum likelihood estimates. In both methods we tried to find an estimate $\hat{\theta}$ in the set Θ such that the distribution $\mathbb{P}_{\hat{\theta}}$ in some sense best describes the data. We didn't make any additional assumptions about the nature of the sample and used only the sample to construct the estimate of θ_0. In the next few lectures we will discuss a different approach to this problem called Bayes estimators. In this approach one would like to incorporate into the estimation process some apriori intuition or theory about the parameter θ_0. The way one describes this apriori intuition is by considering a distribution on the set of parameters Θ or, in other words, one thinks of parameter θ as a random variable. Let $\xi(\theta)$ be a p.d.f. of p.f. of this distribution which is called *prior distribution*. Let us emphasize that $\xi(\theta)$ does not depend on the sample $X_1, \ldots, X_n$, it is chosen apriori, i.e. before we even see the data.

Example. Suppose that the sample has Bernoulli distribution $B(p)$ with p.f.

$$f(x|p) = p^x(1-p)^{1-x} \text{ for } x = 0, 1,$$

where parameter $p \in [0, 1]$. Suppose that we have some intuition that unknown parameter should be somewhere near 0.4. Then $\xi(p)$ shown in figure 9.1 can be a possible choice of a prior distribution that reflects our intution.

$\square$

After we choose prior distribution we observe the sample $X_1, \ldots, X_n$ and we would like to estimate the unknown parameter θ_0 using both the sample and the prior distribution. As a first step we will find what is called the *posterior distribution*

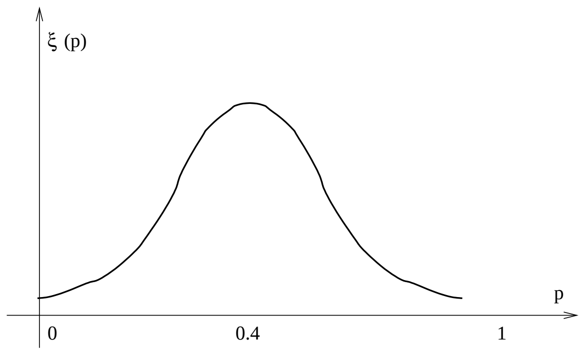

Figure 9.1: Prior distribution.

of θ which is the distribution of θ given $X_1, \ldots, X_n$. This can be done using Bayes theorem.

Total probability and Bayes theorem. If we consider a disjoint sequence of events $A_1, A_2, \ldots$ so that $A_i \cap A_j = \emptyset$ and $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = 1$ then for any event B we have

$$\mathbb{P}(B) = \sum_{i=1}^{\infty} \mathbb{P}(B \cap A_i).$$

Then the Bayes Theorem states the equality obtained by the following simple compuation:

$$\mathbb{P}(A_1|B) = \frac{\mathbb{P}(A_1 \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_1)}{\sum_{i=1}^{\infty} \mathbb{P}(B \cap A_i)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_1)}{\sum_{i=1}^{\infty} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}.$$

We can use Bayes formula to compute

$$\xi(\theta|X_1, \ldots, X_n) - \text{ p.d.f. or p.f. of } \theta \text{ given the sample}$$

if we know

$$f(X_1, \ldots, X_n|\theta) = f(X_1|\theta) \ldots f(X_n|\theta)$$

- p.d.f. or p.f. of the sample given θ, and if we know the p.d.f. or p.f. $\xi(\theta)$ of θ. Posterior distribution of θ can be computed using Bayes formula:

$$\begin{aligned} \xi(\theta|X_1, \ldots, X_n) &= \frac{f(X_1, \ldots, X_n|\theta)\xi(\theta)}{\int_\Theta f(X_1, \ldots, X_n|\theta)\xi(\theta)d\theta} \\ &= \frac{f(X_1|\theta) \ldots f(X_n|\theta)\xi(\theta)}{g(X_1, \ldots, X_n)} \end{aligned}$$

where

$$g(X_1, \ldots, X_n) = \int_\Theta f(X_1|\theta) \ldots f(X_n|\theta)\xi(\theta)d\theta.$$

Example. Very soon we will consider specific choices of prior distributions and we will explicitly compute the posterior distribution but right now let us briefly give an example of how we expect the data and the prior distribution affect the posterior distribution. Assume again that we are in the situation described in the above example when the sample comes from Bernoulli distribution and the prior distribution is shown in figure 9.1 when we expect p_0 to be near 0.4 On the other hand, suppose that the average of the sample is $\bar{X} = 0.7$. This seems to suggest that our intuition was not quite right, especially, if the sample size if large. In this case we expect that posterior distribution will look somewhat like the one shown in figure 9.2 - there will be a balance between the prior intuition and the information contained in the sample. As the sample size increases the maximum of prior distribution will eventually shift closer and closer to $\bar{X} = 0.7$ meaning that we have to discard our intuition if it contradicts the evidence supported by the data.

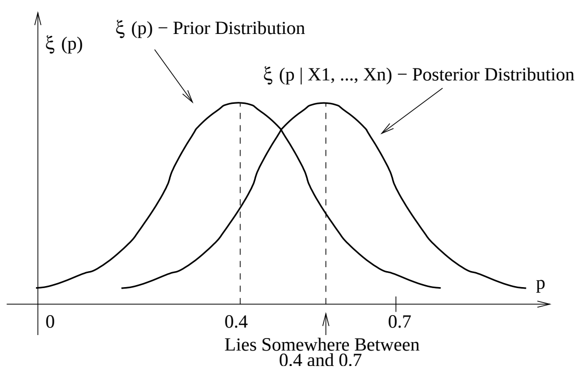

Figure 9.2: Posterior distribution.