COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

STAT 4224/5224

Bayesian Statistics

Dobrin Marchev

Review last lecture: Example 0

A factory uses tools of a particular type. From time-to-time failures in these tools occur and they need to be replaced. The number of such failures in a day has a Poisson distribution with mean 1.25. At the beginning of a particular day there are only five replacement tools in stock. A new delivery of replacements will arrive after four days. If all five spares are used before the new delivery arrives then further replacements cannot be made until the delivery arrives.

Find

(a) the probability that three replacements are required over the next four days.

(b) the expected number of replacements actually made over the next four days.

Joint Distributions

Discrete case:

Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be two countable sample spaces, and X_1 and X_2 be two discrete random variables taking values in these spaces.

The *joint* probability mass (density) function of X_1 and X_2 is defined as

$$p(x_1, x_2) = \mathrm{P}(X_1 = x_1 \cap X_2 = x_2) = \mathrm{P}(X_1 = x_1, X_2 = x_2)$$

The *marginal* pdf of X_1 is then computed as

$$p_1(x_1) = \sum_{x_2 \in \mathcal{X}_2} p(x_1, x_2)$$

The *conditional* pdf of X_2 given $X_1 = x_1$ is computed as

$$p(x_2|x_1) = \frac{p(x_1, x_2)}{p_1(x_1)}$$

Example 1

Suppose the joint distribution of X_1 and X_2 is given below:

		X_2		
		1	2	3
	1	0.32	0.03	0.01
X_1	2	0.06	0.24	0.02
	3	0.02	0.03	0.27

a) Find the two marginal pdfs p_1 and p_2.

b) Find the conditional pdf $p(x_2 \mid X_1 = 2)$.

Example 1 Solution

		X_2			
		1	2	3	p_1
X_1	1	0.32	0.03	0.01	0.36
	2	0.06	0.24	0.02	0.32
	3	0.02	0.03	0.27	0.32
	p_2	0.4	0.3	0.3	

$p(x_2 \mid X_1 = 2)$		
1	2	3
0.06/0.32	0.24/0.32	0.02/0.32

$p(x_2 \mid X_1 = 2)$		
1	2	3
0.1875	0.75	0.0625

Exercise 1

A jar contains 30 red marbles, 50 green marbles and 20 blue marbles. A sample of 15 marbles is selected *with replacement*. Let X be the number of red marbles and Y be the number of blue marbles.

a) What is the joint probability mass function of X and Y ?

b) Repeat part a) but this time if the selection is done *without replacement.*

c) Under which selection scheme is $P(X = 5, Y = 0)$ more likely?

Answers:

a) Multinomial

Joint Distributions

Continuous case:

Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be two uncountable sample spaces, and X_1 and X_2 be two continuous random variables taking values in these spaces. The *joint* probability density function of X_1 and X_2 is such that

$$f(x_1, x_2) \geq 0$$

$$\iint f(x_1, x_2) dx_1 dx_2 = 1$$

The *marginal* pdf of X_1 is then computed as

$$f_1(x_1) = \int f(x_1, x_2) dx_2$$

The *conditional* pdf of X_2 given $X_1 = x_1$ is computed as

$$f(x_2 | x_1) = \frac{f(x_1, x_2)}{f_1(x_1)}$$

Example 2

Let the joint pdf of X_1 and X_2 be:

$$f(x_1, x_2) = 2e^{-(x_1+x_2)}, 0 \leq x_1 \leq x_2 \text{ and } x_2 \geq 0$$

Obtain $f_1(x_1)$.

Example 2 Solution

Let the joint pdf of X_1 and X_2 be:

$$f(x_1, x_2) = 2e^{-(x_1+x_2)}, 0 \leq x_1 \leq x_2 \text{ and } x_2 \geq 0$$

Obtain $f_1(x_1)$.

$$f_1(x_1) = \int_{x_1}^{\infty} 2e^{-(x_1+x_2)}\, dx_2 = 2e^{-x_1} \int_{x_1}^{\infty} e^{-x_2}\, dx_2 = 2e^{-x_1} e^{-x_1}$$

$$= 2e^{-2x_1}, x_1 \geq 0$$

That is, $X_1 \sim \mathrm{Exp}(2)$.

Exercise 2

Let the joint pdf of X_1 and X_2 be:

$$f(x_1, x_2) = \frac{x_1 x_2}{16}, 0 \leq x_1 \leq 2 \text{ and } 0 \leq x_2 \leq 4$$

Find $P(X_2 < X_1^2)$

Answer: 1/3

Joint Distributions

Mixed case:

Let X_1 be discrete and X_2 be continuous random variable.
Then the joint probability density function $f(x_1, x_2)$ of X_1 and X_2 is handled by adding over x_1 and integrating over x_2.

Very often, in Bayesian analysis, the mixed case arises from the multiplication law:

$$f(x_1, x_2) = p_1(x_1) f(x_2 \mid x_1)$$

or

$$f(x_1, x_2) = f_2(x_2) p(x_1 \mid x_2)$$

Example 3

Let the joint pdf of X_1 and X_2 be:

$$f(x_1, x_2) = \frac{x_1 x_2^{x_1 - 1}}{3}, x_1 = 1, 2, 3, \text{and } 0 < x_2 < 1$$

Verify that this is a valid pdf.

Example 3 Solution

Let the joint pdf of X_1 and X_2 be:

$$f(x_1, x_2) = \frac{x_1 x_2^{x_1 - 1}}{3}, x_1 = 1, 2, 3, \text{and } 0 < x_2 < 1$$

Verify that this is a valid pdf.

We need to verify that

$$\sum_{x_1=1}^{3} \int_0^1 \frac{x_1 x_2^{x_1 - 1}}{3} dx_2 = 1$$

But $\int_0^1 \frac{x_1 x_2^{x_1 - 1}}{3} dx_2 = \frac{1}{3}$ and $\sum_{x_1=1}^{3} \frac{1}{3} = 1$, so it is a valid pdf.

Exercise 3

Let the joint pdf of X_1 and X_2 be:

$$f(x_1, x_2) = \frac{x_1 x_2^{x_1 - 1}}{3}, x_1 = 1, 2, 3, \text{and } 0 < x_2 < 1$$

Find $P\left(X_1 \geq 2, X_2 \geq \frac{1}{2}\right)$

Answer: 0.5417

Connection with Bayesian Analysis

One example of a mixture of discrete and continuous random variables will be presented next week as the Beta-Binomial model, where:

θ = proportion of people in a large population who have a certain characteristic (continuous variable between 0 and 1).

X = number of people in a random sample from the population who have the characteristic (discrete Binomial variable).

Bayesian estimation of θ derives from the calculation of $f(\theta|x)$, where x is the observed value of X. This calculation first requires that we have a joint mixed density $f(x, \theta)$ obtained by multiplying the prior and the likelihood: $f(x, \theta) = \pi(\theta)p(x \mid \theta)$

Then

$$f(\theta|x) = \frac{f(\theta, x)}{p(x)} = \frac{\pi(\theta)p(x|\theta)}{p(x)} = \frac{\pi(\theta)p(x|\theta)}{\int \pi(\theta)p(x|\theta)d\theta}$$

Independence of Random Variables

Definition: Suppose that $X_1, \ldots, X_n$ are random variables. We say that $X_1, \ldots, X_n$ are independent if for every collection of events $A_1, \ldots, A_n$ we have

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \times \ldots \times P(X_n \in A_n)$$

However, in Bayesian analysis it is more important the the random variables are conditionally independent given the parameter value.

Definition: Suppose that $X_1, \ldots, X_n$ are random variables and θ is a parameter. We say that $X_1, \ldots, X_n$ are conditionally independent given θ if for every collection of events $A_1, \ldots, A_n$ we have

$$P(X_1 \in A_1, \ldots, X_n \in A_n \mid \theta) = P(X_1 \in A_1 \mid \theta) \times \ldots \times P(X_n \in A_n \mid \theta)$$

Example 1 (Continued)

Suppose the joint distribution of X_1 and X_2 is given below:

		X_2		
		1	2	3
X_1	1	0.32	0.03	0.01
	2	0.06	0.24	0.02
	3	0.02	0.03	0.27

c) Are X_1 and X_2 independent random variables?

Example 1 Solution

		X_2			
		1	2	3	p_1
X_1	1	0.32	0.03	0.01	0.36
	2	0.06	0.24	0.02	0.32
	3	0.02	0.03	0.27	0.32
	p_2	0.4	0.3	0.3	

If X_1 and X_2 independent random variables we must have that, among other equalities, $P(X_1 = 1, X_2 = 1) = P(X_1 = 1)\, P(X_2 = 1)$.
But
$P(X_1 = 1, X_2 = 1) = 0.32 \neq 0.144 = 0.36(0.4) = P(X_1 = 1)\, P(X_2 = 1)$.
Therefore, X_1 and X_2 are *not* independent.

Exercise 2 (Continued)

Let the joint pdf of X_1 and X_2 be:

$$f(x_1, x_2) = \frac{x_1 x_2}{16}, 0 \leq x_1 \leq 2 \text{ and } 0 \leq x_2 \leq 4$$

Obtain $f_1(x_1)$ and $f_2(x_2)$ and verify that $f(x_1, x_2) = f_1(x_1) f_2(x_2)$, confirming that X_1 and X_2 are independent.

Connection with Bayesian Analysis

In Bayesian analysis, under the assumption of conditional independence the joint density (likelihood) is given by:

$$f(x_1, \dots, x_n|\theta) = f_1(x_1|\theta) \times \dots \times f_n(x_n|\theta) = \prod_{i=1}^{n} f_i(x_i|\theta)$$

Furthermore, we will very often assume that $X_1, \dots, X_n$ are generated by a common process, that is, they have the same marginal density $f(x \mid \theta)$, so the likelihood is then

$$f(x_1, \dots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

In this case we say that $X_1, \dots, X_n$ are *conditionally independent and identically distributed* or simply i.i.d.

Example 4

Suppose that

$$X_1, X_2, \dots, X_n \,|\theta \sim Bernoulli(\theta)$$

Obtain the likelihood function $\mathrm{L}(\theta) = f(x_1, \dots, x_n|\theta)$

Example 4 Solution

$$X_1, X_2, \ldots, X_n | \theta \sim Bernoulli(\theta)$$

Find the likelihood.

Note that $f(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$

Therefore,

$$\mathrm{L}(\theta) = f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} [\theta^{x_i}(1-\theta)^{1-x_i}]$$

$$= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{\sum_{i=1}^{n}(1-x_i)}$$

$$= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

Exercise 4

Suppose that

$$X_i \mid \theta \sim \mathrm{Exp}\,(\theta)$$

and we have observed

$$(x_1, x_2, x_3, x_4) = (1.23, 3.32, 1.98, 2.12).$$

Obtain and plot the likelihood function.

Exchangeability

Definition: Let $f(x_1, \dots, x_n)$ be the joint density of $X_1, \dots, X_n$. If

$$f(x_1, \dots, x_n) = f\left(x_{\pi_1}, \dots, x_{\pi_n}\right)$$

for any permutation π of $\{1, 2, \dots n\}$, then we say that $X_1, \dots, X_n$ are *exchangeable.*

Roughly speaking, $X_1, \dots, X_n$. are exchangeable if the subscript labels convey no information about the outcomes.

Note: Independence is a stronger property, and it implies exchangeability, but not vice versa.

Example 5

Suppose you have an urn containing 1 red balls and 2 white balls. Draw out balls, one at a time and *without* replacement, and note the color.

Define

$$X_i = \begin{cases} 1, \text{if the } i^{\text{th}} \text{ ball is red} \\ 0, \qquad\qquad\qquad\quad \text{o/w} \end{cases}$$

Prove that X_1, X_2 and X_3 are exchangeable, but not independent.

Example 5 Solution

The only nonzero propagability outcomes are (1, 0, 0), (0, 1, 0) and (0, 0, 1).

$$P(X_1 = 1, X_2 = 0, X_3 = 0) = \frac{1}{3} \times 1 \times 1 = \frac{1}{3}$$
$$P(X_1 = 0, X_2 = 1, X_3 = 0) = \frac{2}{3} \times \frac{1}{2} \times 1 = \frac{1}{3}$$
$$P(X_1 = 0, X_2 = 0, X_3 = 1) = \frac{2}{3} \times \frac{1}{2} \times 1 = \frac{1}{3}$$

Since these are all the same, the random variables X_1, X_2, and X_3 are exchangeable.

Exercise 5

Prove that X_1, X_2, and X_3 in Example 5 are dependent.

Connection with Bayesian Analysis

Suppose we consider the Bernoulli example as Bayesian:

$$X_1, X_2, \ldots, X_n \mid \theta \sim Bernoulli(\theta)$$
$$\theta \sim \pi(\theta)$$

where $\pi(\theta)$ represents our prior belief about the parameter.

Then the joint distribution of the data and the parameter is:

$$\theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}\pi(\theta)$$

and the marginal distribution of the data is:

$$p(x_1, \ldots, x_n) = \int_0^1 \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}\pi(\theta)d\theta$$

Note that no matter what permutation of 0 and 1 you insert, this marginal distribution is the same.

Result: If $X_1, \ldots, X_n$ are conditionally i.i.d. given θ, then marginally (unconditionally on θ), $X_1, \ldots, X_n$ are exchangeable.

Summary

- iid implies exchangeability.
- Exchangeability implies identically distributed.
- In Bayesian analysis, the data are marginally exchangeable.