Lecture 2

Axioms of Probability

Kolmogorov's Axioms of Probability:

The textbook discusses the idea of belief functions.

Partitions and Bayes Rule

A collection of sets $\{H_1, ..., H_K\}$ is a **partition** of another set $\mathcal{H}$ if

Let $\mathcal{H}$ be a Bozemanite's favorite ski hill. Partitions include:

Let $\mathcal{H}$ be the number of stats courses taken by people in this room. Partitions include:

Suppose $\{H_1, ..., H_K\}$ is a partition of $\mathcal{H}$, $Pr(\mathcal{H}) = 1$, and E is some specific event. The axioms of probability imply the following statements:

1. **Rule of Total Probability:**

2. **Rule of Marginal Probability:**

3. **Bayes rule:**

Assume a sample of MSU students are polled on their skiing behavior. Let $\{H_1, H_2, H_3, H_4\}$ be the events that a randomly selected student in this sample is in, the first quartile, second quartile, third quartile and 4th quartile in terms of number of hours spent skiing.

Then $\{Pr(H_1), Pr(H_2), Pr(H_3), Pr(H_4)\} = \{.25, .25, .25, .25\}$.

Let E be the event that a person has a GPA greater than 3.0, where $\{Pr(E|H_1), Pr(E|H_2), Pr(E|H_3), Pr(E|H_4)\} = \{.40, .71, .55, .07\}$.

Now compute the probability that a student with a GPA greater than 3.0 falls in each quartile for hours spent skiing : $\{Pr(H_1|E), Pr(H_2|E), Pr(H_3|E), Pr(H_4|E)\}$

Independence

Two events F and G are conditionally independent given H if

If F and G are conditionally independent given H then

What is the relationship between $Pr(F|H)$ and $Pr(F|H \cap G)$ as well as $Pr(F|H)$ and $Pr(F|H \cap I)$? - $F = ${ you draw the jack of hearts } - $G = ${ a mind reader claims you drew the jack of hearts } - $H = ${ the mind reader has extrasensory perception } - $I = ${ Andy is the mind reader }

Random Variables

In Bayesian inference a random variable is defined as an unknown numerical quantity about which we make probability statements.

Discrete Random Variables

Let Y be a random variable and let $\mathcal{Y}$ be the set of all possible values of Y.

The event that the outcome Y of our study has the value y is expressed as $\{Y = y\}$.

This function (known as the probability distribution function (pdf)) $p(y)$ has the following properties.

Example 1. Binomial Distribution

Let $\mathcal{Y} = \{0, 1, 2, ..n\}$ for some positive integer n. Then $Y \in \mathcal{Y}$ has a binomial distribution with probability θ if

Example 2. Poisson Distribution

Let $\mathcal{Y} = \{0, 1, 2...\}$. Then $Y \in \mathcal{Y}$ has a Poisson distribution with mean θ if

Continuous Random Variables

Suppose that the sample space $\mathcal{Y}$ is $\mathbb{R}$, then $Pr(Y \leq 5) \neq \sum_{y \leq 5} p(y)$ as this sum does not make sense. Rather define the cumulative distribution function (cdf) $F(y) = Pr(Y \leq y)$.

The cdf has the following properties:

Using the CDF, probabilities of events can be derived as:

If F is continuous, then Y is a continuous random variable. Then $F(a) = \int_{-\infty}^{a} p(y)dy$.

Example. Normal distribution.

Let $\mathcal{Y} = (-\infty, \infty)$ with mean μ and variance σ^2. Then y follows a normal distribution if

Moments of Distributions

The mean or expectation of an unknown quantity Y is given by

The variance is a measure of the spread of the distribution.

$$
\begin{aligned}
Var[Y] &= E[(Y - E[Y])^2] \\
&= E[Y^2 - 2YE[Y] + E[Y]^2] \\
&= E[Y^2] - 2E[Y]^2 + E[Y]^2 \\
&= E[Y^2] - E[Y]^2
\end{aligned}
$$

If $Y \sim$ Binomial(n, p), then

if $Y \sim$ Poisson(μ), then

if $Y \sim$ Normal(μ, σ^2),

Joint Distributions

Let Y_1, Y_2 be random variables, then the joint pdf or joint density can be written as

The marginal density of Y_1 can be computed from the joint density:

Note this is for discrete random variables, but a similar derivation holds for continuous.

The conditional density of Y_2 given $\{Y_1 = y_1\}$ can be computed from the joint density and the marginal density.

$$
\begin{aligned}
p_{Y_2|Y_1}(y_2|y_1) &= \frac{Pr(\{Y_1 = y_1\} \cap \{Y_2 = y_2\})}{Pr(Y_1 = y_1)} \\
&= \frac{p_{Y_1,Y_2}(y_1, y_2}{p_{Y_1}(y_1)}
\end{aligned}
$$

Note the subscripts are often dropped, so $p_{Y_1,Y_2}(y_1, y_2) = p(y_1, y_2)$, ect...

Independent Random Variables and Exchangeability

Suppose $Y_1, ..., Y_n$ are random variables and that θ is a parameter corresponding to the generation of the random variables. Then $Y_1, ..., Y_n$ are conditionally independent given θ if

where $\{A_1, ..., A_n\}$ are sets.

Then the joint distribution can be factored as

If the random variables come from the same distribution then they are conditionally independent and identically distributed, which is noted $Y_1, ..., Y_n|\theta \sim i.i.d. p(y|\theta$ and

$$p(y_1, ..., y_n|\theta) = p_{Y_1}(y1|\theta) \times ... \times p_{Y_n}(y_n|\theta) = \prod_{i=1}^{n} p(y_i|\theta).$$

Exchangeability

Let $p(y_1, ...y_n)$ be the joint density of $Y_1, ..., Y_n$. If $p(y_1, ..., y_n) = p(y_{\pi_1}, ..., y_{\pi_n})$ for all permutations π of {1, 2, . . . ,n}, then $Y_1, ..., Y_n$ are exchangeable.

Assume data has been collected on apartment vacancies in Bozeman. Let $y_i = 1$ if an *affordable* room is available. Do we expect $p(y_1 = 0, y_2 = 0, y_3 = 0, y_4 = 1) = p(y_1 = 1, y_2 = 0, y_3 = 0, y_4 = 0)$? If so the data are exchangeable.

Let $\theta \sim p(\theta)$ and if $Y_1, ..., Y_n$ are conditionally i.i.d. given θ, then marginally (unconditionally on θ) $Y_1, ..., Y_n$ are exchangeable. Proof omitted, see textbook for details.