Lecture 2 - Key

Axioms of Probability

Kolmogorov's Axioms of Probability:

- $0 = Pr(not\ H|H) \leq Pr(F|H) \leq Pr(H|H) = 1$

- $Pr(F \cup G|H) = Pr(F|H) + Pr(G|H)\ if\ F \cap G = \emptyset$

- $Pr(F \cap G|H) = Pr(G|H)Pr(F|G \cap H)$

The textbook discusses the idea of belief functions. *The important thing to note here is that a probability function can be used to express beliefs in a principled manner.*

Partitions and Bayes Rule

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

- *the events are disjoint,* $H_i \cap H_j = \emptyset$ *for* $i \neq j$ *and*

- *the union of the sets is* $\mathcal{H}$, $\cup_{k=1}^{K} H_k = \mathcal{H}$.

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

- *{Big Sky, Bridger Bowl, other}*

- *{Any Montana Ski Hill, Any Colorado Ski Hill, other}*

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

- *{0, 1, 2, . . . }*

- *{0 - 3, 4-6, 7-10, 10+}*

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:** $\sum_{k=1}^{K} Pr(H_k) = 1$

2. **Rule of Marginal Probability:**

$$\begin{aligned} Pr(E) &= \sum_{k=1}^{K} Pr(E \cap H_k) \\ &= \sum_{k=1}^{K} Pr(E|H_k)Pr(H_k) \end{aligned}$$

3. **Bayes rule:**

$$\begin{aligned} Pr(H_j|E) &= \frac{Pr(E|H_j)Pr(H_j)}{Pr(E)} \\ &= \frac{Pr(E|H_j)Pr(H_j)}{\sum_{k=1}^{K} Pr(E|H_k)Pr(H_k)} \end{aligned}$$

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 : $\{Pr(H_1|E), Pr(H_2|E), Pr(H_3|E), Pr(H_4|E)\}$

$$\begin{aligned} Pr(H_1|E) &= \frac{Pr(E|H_1)Pr(H_1)}{\sum_{k=1}^{4} Pr(E|H_k)Pr(H_k)} \\ &= \frac{Pr(E|H_1)}{Pr(E|H_1) + Pr(E|H_2) + Pr(E|H_3) + Pr(E|H_4)} \\ &= \frac{.40}{.40 + .71 + .55 + .07} = \frac{.4}{1.73} = .23 \end{aligned}$$

Similarly, $Pr(H_2|E) = .41$*,* $Pr(H_3|E) = .32$*, and* $Pr(H_4|E) = .04$*.*

Independence

Two events F and G are conditionally independent given H if $Pr(F \cap G|H) = Pr(F|H)Pr(G|H)$

If F and G are conditionally independent given H then $Pr(F|H \cap G) = Pr(F|H)$

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 = $\{Andyisthemindreader\}$*Pr(F|H) = 1/52$ and $Pr(F|G \cap H) > 1/52$ $Pr(F|H) = 1/52$ and $Pr(F|G \cap I) = 1/52$*

Random Variables

In Bayesian inference a random variable is defined as an unknown numerical quantity about which we make probability statements. *For example, the quantitative outcome of a study is performed. Additionally, a fixed but unknown population parameter is also a random variable.*

Discrete Random Variables

Let Y be a random variable and let $\mathcal{Y}$ be the set of all possible values of Y. Y *is discrete if the set of possible outcomes is countable, meaning that* $\mathcal{Y}$ *can be expressed as* $\mathcal{Y} = \{y_1, y_2, ...\}$*.*

The event that the outcome Y of our study has the value y is expressed as $\{Y = y\}$. *For each* $y \in \mathcal{Y}$*, our shorthand notation for* $Pr(Y = y)$ *will be* $p(y)$*.*

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

- $0 \leq p(y) \leq 1$ *for all* $y \in \mathcal{Y}$

- $\sum_{y \in \mathcal{Y}} p(y) = 1$

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

$$*Pr(Y = y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}.*$$

Example 2. Poisson Distribution

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

$$*Pr(Y = y|\theta) = \theta^y \exp(-\theta)/y!*$$

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:

- $F(\infty) = 1$
- $F(-\infty) = 0$
- $F(b) \leq F(a)$ *if* $b < a$.

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

- $Pr(Y > a) = 1 - F(a)$
- $Pr(a < Y < b) = F(b) - F(a)$

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

$$*p(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}*$$

Moments of Distributions

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

$$\begin{aligned} *E[Y] &= \sum_{y \in \mathcal{Y}} yp(y) \text{ if Y is discrete and } * \\ *E[Y] &= \int_{y \in \mathcal{Y}} yp(y) \text{ if Y is continuous.} * \end{aligned}$$

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 $E[Y] = np$ *and* $Var[Y] = np(1 - p)$.

if $Y \sim$ Poisson(μ), then $E[Y] = \mu$ *and* $Var[Y] = \mu$.

if $Y \sim$ Normal(μ, σ^2), *then* $E[Y] = \mu$ *and* $Var[Y] = \sigma^2$.

Joint Distributions

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

$$*P_{Y_1,Y_2}(y_1, y_2) = Pr(\{Y_1 = y_1\} \cap \{Y_2 = y_2\}), \text{ for } y_1 \in \mathcal{Y}_1, y_2 \in \mathcal{Y}_2 *$$

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

$$\begin{aligned} p_{Y_1}(y_1) &= Pr(Y_1 = y_1) \\ * &= \sum_{y2 \in \mathcal{Y}_2} Pr(\{Y_1 = y_1\} \cap \{Y_2 = y_2\}) \\ * &= \sum_{y2 \in \mathcal{Y}_2} p_{Y_1,Y_2}(y_1, y_2).* \end{aligned}$$

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

$$*Pr(Y_1 \in A_1, ..., Y_n \in A_n|\theta) = Pr(Y_1 \in A_1) \times ... \times Pr(Y_n \in A_n)*$$

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

Then the joint distribution can be factored as

$$*p(y_1, ..., y_n|\theta) = p_{Y_1}(y1|\theta) \times ... \times p_{Y_n}(y_n|\theta).*$$

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.