Probability Recap

David Dalpiaz

STAT 430, Fall 2017

Questions? Comments? Concerns?

Probability Rules

Complement Rule:

$$P[A^c] = 1 - P[A]$$

Conditional Probability:

$$P[A \mid B] = \frac{P[A \cap B]}{P[B]}$$

Probability Rules

Multiplication Rule:

$$P[A \cap B] = P[B] \cdot P[A \mid B] \cdot$$

For a number of events $E_1, E_2, \ldots E_n$, the multiplication rule can be expanded into the **chain rule**:

$$P\left[\bigcap_{i=1}^{n} E_i\right] = P[E_1] \cdot P[E_2 \mid E_1] \cdot P[E_3 \mid E_1 \cap E_2] \cdots P\left[E_n \mid \bigcap_{i=1}^{n-1} E_i\right]$$

Bayes' Theorem

Define a **partition** of a sample space Ω to be $A_1, A_2, \ldots, A_n$ such that

$$A_i \cap A_j = \emptyset$$

for all $i \neq j$, and

$$\bigcup_{i=1}^{n} A_i = \Omega.$$

Let $A_1, A_2, \ldots, A_n$ form a partition of some sample space. Then for an event B we have **Bayes' Theorem**:

$$P[A_i|B] = \frac{P[A_i]P[B|A_i]}{P[B]} = \frac{P[A_i]P[B|A_i]}{\sum_{i=1}^{n} P[A_i]P[B|A_i]}$$

Indepedence

Two events A and B are said to be **independent** if they satisfy

$$P[A \cap B] = P[A] \cdot P[B]$$

A collection of events $E_1, E_2, \ldots E_n$ is said to be independent if

$$P\left[\bigcap_{i \in S} E_i\right] = \prod_{i \in S} P[E_i]$$

for every subset S of $\{1, 2, \ldots n\}$.

If this is the case, then the chain rule is greatly simplified to:

$$P\left[\bigcap_{i=1}^{n} E_i\right] = \prod_{i=1}^{n} P[E_i]$$

Distributions

We often talk about the **distribution** of a random variable, which can be thought of as:

distribution = list of possible **values** + associated **probabilities**

Discrete Random Variables

The distribution of a discrete random variable X is most often specified by a list of possible values and a probability **mass** function, $p(x)$. The mass function directly gives probabilities, that is,

$$p(x) = p_X(x) = P[X = x].$$

Binomial Distribution

$$p(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n, \; n \in \mathbb{N}, \; 0 < p < 1.$$

- The function $p(x|n,p)$ is the mass function.
  - It is a function of x, the possible values of the random variable X.
  - It is conditional on the **parameters** n and p.
- $x = 0, 1, \ldots, n$ indicates the **sample space**.
- $n \in \mathbb{N}$ and $0 < p < 1$ specify the **parameter space**.

$$X \sim \text{bin}(n,p).$$

Continuous Random Variables

The distribution of a continuous random variable X is most often specified by a set of possible values and a probability **density** function, $f(x)$.

The probability of the event $a < X < b$ is calculated as

$$P[a < X < b] = \int_a^b f(x)dx.$$

Note that densities are **not** probabilities.

Normal Distribution

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],$$

$$\infty < x < \infty,\ -\infty < \mu < \infty,\ \sigma > 0.$$

- The function $f(x|\mu, \sigma^2)$ is the density function.
  - It is a function of x, the possible values of the random variable X.
  - It is conditional on the **paramters** μ and σ^2.
  - $-\infty < x < \infty$ indicates the sample space.
  - $-\infty < \mu < \infty$ and $\sigma > 0$ specify the parameter space.

$$X \sim N(\mu, \sigma^2)$$

Expectations

For discrete random variables, we define the **expectation** of the function of a random variable X as follows.

$$\mathbb{E}[g(X)] \triangleq \sum_{x} g(x)p(x)$$

For continuous random variables we have a similar definition.

$$\mathbb{E}[g(X)] \triangleq \int g(x)f(x)dx$$

For specific functions g, expectations are given names.

Mean

The **mean** of a random variable X is given by

$$\mu_X = \text{mean}[X] \triangleq \mathbb{E}[X].$$

So for a discrete random variable, we would have

$$\text{mean}[X] = \sum_x x \cdot p(x)$$

For a continuous random variable we would simply replace the sum by an integral.

Variance

The **variance** of a random variable X is given by

$$\sigma_X^2 = \mathrm{var}[X] \triangleq \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

The **standard deviation of a random variable X is given by

$$\sigma_X = \mathrm{sd}[X] \triangleq \sqrt{\sigma_X^2} = \sqrt{\mathrm{var}[X]}.$$

The **covariance** or random variables X and Y is given by

$$\mathrm{cov}[X, Y] \triangleq \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Likelihood

Consider n iid random variables $X_1, X_2, \ldots X_n$. We can then write their **likelihood** as

$$\mathcal{L}(\theta \mid x_1, x_2, \ldots x_n) = \prod_{i=i}^{n} f(x_i; \theta)$$

where $f(x_i; \theta)$ is the density (or mass) function of random variable X_i evaluated at x_i with parameter θ.

Whereas a probability is a function of a possible observed value given a particular parameter value, a likelihood is the opposite. It is a function of a possible parameter value given observed data.