Notes

Abhinav Patil

Jan. 29, 2019

1 Syllabus

Discussed syllabus (read it).

2 Random Variables

Let X be a random variable (or a **data-generating process**). Then x (data) refers to a **realization** of X.

We define the **support** of X, denoted $\text{Supp}[X]$, as the set of all possible unique/different values of x.

There are two types of random variables: **discrete** and **continuous**.

2.1 Discrete Random Variables

For discrete random variables, $|\text{Supp}(X)| \leq |\mathbb{N}|$. That is, the cardinality of the support of X is at most the cardinality of the natural numbers. In other words, the number of unique values of x is either finite or countably infinite.

2.1.1 Probability Mass Function (PMF)

We define the **probability mass function**, $p(x)$, of a discrete random variable as follows:

$$p(x) := P(X = x)$$

Clearly, $p(x)$ is a probability. Then what is its domain and range?

$$p : \text{Supp}[x] \longrightarrow (0, 1]$$

Why do we exclude zero? Recall the definition of the support: it only includes *what is possible*. Thus anything in the support has non-zero probability. However, we include 1 because a random variable X could produce the same realization x 100% of the time.

2.1.2 Cumulative Distribution Function (CDF)

We define the **cumulative distribution function**, $F(x)$ as follows:

$$\begin{aligned} F(x) :&= P(X \leq x) \\ &= \sum_{y \in \text{Supp}[X] \text{ and } y \leq x} p(y) \end{aligned}$$

Note that:

$$\sum_{x \in \text{Supp}[X]} p(x) = 1$$

2.2 Continuous Random Variables

For continuous random variables, $|\text{Supp}(X)| = |\mathbb{R}|$. That is, the cardinality of the support of X is uncountably infinite and equal to the cardinality of the reals.

2.2.1 Probability Density Function

We define the **probability density function**, $f(x)$, of a continuous random variable as follows:

$$f(x) := F'(x) = \frac{\mathrm{d}}{\mathrm{d}x}[F]$$

Observe that:

$$\begin{aligned} p(x \in [a, b]) &= p(x \leq b) - p(x \leq a) \\ &= F(b) - F(a) \\ &= \int_a^b f(x) dx \end{aligned}$$

n.b.: The following is abuse of notation; we should use limits.

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx &= F(\infty) - F(-\infty)\\ &= p(x \leq \infty) - p(x \leq -\infty)\\ &= p(x \in (-\infty, \infty))\\ &= 1\end{aligned}$$

Domain and range:

$$f : \text{Supp}[x] \longrightarrow [0, \infty)$$

Support: $\text{Supp}(x) = \{x : f(x) > 0\}$

n.b.: $f(x)$ IS NOT the same as $P(X = x) = p(x)$. $f(x)$ is not a probability; however, integrating over $f(x)$ does yield a probability.

3 Models & Parameters

We can identify random variables by their cdf/pmf if they are discrete and their cdf/pdf if they are continuous. Examples:

Discrete:

$$\begin{aligned}&X \sim \text{Bernoulli}(p) := p^x(1-p)^{1-x} && x \in \text{Supp}[X] = \{0, 1\}\\ &X \sim \text{Binomial}(n, p) := \binom{n}{x} p^x(1-p)^{n-x} && x \in \text{Supp}[X] = \{0, 1, \ldots, n\}\end{aligned}$$

Here, means "distributed as." Note that $p(x) = p^x(1-p)^{1-x}$

Continuous:

$$\begin{aligned}&X \sim \text{Exp}(\lambda) := \lambda e^{-\lambda x} && x \in (0, \infty)\\ &X \sim \text{N}(\mu, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}}(x-\mu)^2 && x \in \mathbb{R}\end{aligned}$$

Exp = exponential; N = normal

3.1 Bernoulli

$P(X = 1) = p(1) = p$ and $P(X = 0) = p(0) = 1 - p$

p is a **parameter**, which we can think of as a "tuning knob" that controls how often 0s and 1s are realized.

Strictly speaking, $p \in [0, 1]$. However, we call it a **degenerate random variable** when $p = 0$ or $p = 1$. We will primarily deal with non-degenerate cases: $p \in (0, 1)$. This is the **parameter space**, the set of values of the parameter that yield non-degenerate random variables.

3.2 Unknown parameters

Let θ denote an unknown parameter, $\overrightarrow{\theta}$ denote multiple unknown parameters, and Θ the parameter space. Then consider the following one-parameter models:

$$
\begin{aligned}
X &\sim \text{Bernoulli}(\theta) = \theta^x (1-\theta)^{1-x} \\
X &\sim \text{Binomial}(n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}
\end{aligned}
$$

In this binomial model, n is known while θ is unknown. We could also deal with a two-parameter binomial model:

$$
X \sim \text{Binomial}(\theta_2, \theta_1) = \binom{\theta_2}{x} \theta^x (1-\theta_1)^{\theta_2 - x}
$$

However in this course we will not be dealing with this.

Similarly, for continuous r.v.'s:

$$
\begin{aligned}
X &\sim \text{Exp}(\theta) := \theta e^{-\theta x} \\
X &\sim \text{N}(\theta, \sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}} (x-\theta)^2
\end{aligned}
$$

3.3 Parametric Models

n.b.: for the remainder of the course, $p(x) = f(x)$.

We define a **parametric model** (set of all possible models of a given type):

$$\mathcal{F} := \{p(x;\theta\} : \theta \in \Theta\}, \text{s.t. } \dim[\Theta] < \infty$$

Notes on notation: "dim" refers to "dimensionality." $p(x;\theta)$ means "probability of x with θ known."

For example, with Bernoulli:

$$\mathcal{F}_{\text{Bern}} = \{\theta^x(1-\theta)^{1-x} : \theta \in (0,1)\}$$

This is the set of all possible Bernoulli models.

3.4 Joint Mass/Density Function

The **joint mass/density function** ask "what is the probability or density of $x_1, x_2, x_3, \dots, x_n$, knowing θ?". We write this as:

$$P(x_1, x_2, x_3, \dots, x_n; \theta)$$

Notation: $\overset{iid}{\sim}$ means "**independent and identically distributed**."

If $x_1, x_2, x_3, \dots, x_n$ are independent and identically distributed, then

$$\begin{aligned} P(x_1, x_2, x_3, \dots, x_n; \theta) &= \Pi_{i=1}^n p_i(x_i; \theta) \\ &= \Pi_{i=1}^n p(x_i; \theta) \end{aligned}$$

4 Real world vs. this class

In the real world, you observe data, e.g. $x =< 0, 0, 1, 0, 1, 0 >$. Thus $n = 6$ (number of observations). Then we have to make an assumption in picking a parametric model (an $\mathcal{F}$). What determines how to pick this model is beyond the scope of this class; here we will always be given a parametric model to work with. Then, given $\mathcal{F}$, we will be **inferring** θ from data using a Bayesian perspective.