1 Distribuciones condicionales

Before, we wrote

$$P(X \in S)$$

Now, we will write

$$P(X \in S|A) \coloneqq \frac{P(\{w : X(w) \in S\} \cap A)}{P(A)}$$

En general,

$$X = (X_1, \ldots, X_d)$$

$$P(X_i \in A|X_j \in B) = \frac{P(X_i) \in A, X_j \in B}{P(X_j \in B)}$$

1.1 Independence of random variable

X, Y are independent random variables random variables if

$$\begin{aligned} \forall A, B : P(x \in A|y \in B) = P(x \in A) \\ \forall A, B : P(y \in B|x \in A) = P(y \in B) \\ \forall A, B : P(x \in A, y \in B) = P(x \in A)P(y \in B) \end{aligned}$$

Definition 1: We say that

$$X \sim Y$$

if the distribution of X equals the distribution of Y.

1.2 Recap

Definition 2: Let X be a discrete random variable. We define de *expected value* of X as

$$E[X] = \sum x \cdot P(X = x)$$

We note that

$$\sum(x - E[X]) \cdot P(X = x) = \sum x \cdot P(X = x) - E[X] \sum P(X = x) = 0$$

Definition 3: For any $g : X \to Y$,

$$E[g(X)] = \sum g(x)P(x = X)$$

Theorem 1: E is linear.

$$E(\alpha X_1 + \beta X_2) = \alpha E[X_1] + \beta E[X_2]$$

Proof.

$$\begin{aligned}
E(\alpha X_1 + \beta X_2) &= \sum_{(x_1, x_2)} (\alpha x_1 + \beta x_2) P(X = (x_1, x_2)) \\
&= \sum_{x_1} \sum_{x_2} (\alpha x_1 + \beta x_2) P(x_1 = X_1, x_2 = X_2) \\
&= \sum_{x_1} \sum_{x_2} \alpha x_1 P(x_1 = X_1, x_2 = X_2) + \sum_{x_1} \sum_{x_2} \beta x_2 P(x_1 = X_1, x_2 = X_2) \\
&= \sum_{x_1} \alpha x_1 P(x_1 = X_1) + \sum_{x_2} \beta x_2 P(x_2 = X_2)
\end{aligned}$$

□

Theorem 2: If X_1, X_2 are independent random variables, then

$$\mathbf{E}[X_1 X_2] = \mathbf{E}[X_1]\,\mathbf{E}[X_2].$$

Proof.

$$\begin{aligned}
\mathbf{E}[X_1 X_2] &= \sum_{x_1} \sum_{x_2} x_1 x_2 P(X_1 = x_1, X_2 = x_2) \\
&= \sum_{x_1} x_1 P(X_1 = x_1) \sum_{x_2} x_2 P(X_2 = x_2) \\
&= \mathbf{E}[X_1]\,\mathbf{E}[X_2]
\end{aligned}$$

□

THe total probability law tells us that

$$\begin{aligned}
\mathbf{E}[X] &= \sum \mathbf{E}[X|Y = y] P(Y = y) \\
&= \sum x P(X = x|Y = y)
\end{aligned}$$

Definition 4: For $X \in \mathbb{R}$,

$$Var(X) = \mathbf{E}[(X - \mathbf{E}[X])^2]$$

Remark 1:

$$\begin{aligned}\mathbf{E}[(X-\mathbf{E}\,X)^2] &= \mathbf{E}[X^2-2X\,\mathbf{E}[X]+\mathbf{E}[X]^2]\\ &= \mathbf{E}[X^2]-\mathbf{E}[2X\,\mathbf{E}[X]]+\mathbf{E}[\mathbf{E}[X]^2]\\ &= \mathbf{E}[X^2]-2E[X]^2+E[X]^2\\ &= \mathbf{E}[X^2]-\mathbf{E}[X]^2.\end{aligned}$$

Theorem 3: For any random variable $X \in \mathbb{R}^n$,

$$\mathbf{Var}(X) \geq 0$$

$$\mathbf{Var}(aX) = a^2\,\mathbf{Var}(X),$$

Given an event A, how can we quantify the level of surprise this event may cause?

Definition 5 (Entropy)**:** Given an event A, we define the its entropy as

$$\mathcal{E}(A) := -\sum \log(P(X=x))P(X=x)$$

Theorem 4: For any event A , $\mathcal{E}(A) \geq 0$.

Problem 1: Let X be a random variable. Define $Y \sim X$, independent of X. Find $P(X \neq Y)$.

Proof. We first notice that

$$\begin{aligned}P(X=Y) &= \sum_y P(X=Y|Y=y)P(Y=y)\\ &= \sum_y P(X=y|Y=y)P(Y=y)\\ &= \sum_y P(X=y)P(Y=y)\\ &= \sum_y P(X=y)^2\end{aligned}$$

therefore,

$$P(X\neq Y) = 1-P(X=Y) = 1-\sum_y P(X=y)^2 = \mathcal{E}_{gini}(X)$$

□

1.3 Distribution examples

Example 1 (Bernoulli)**:** The *Bernoulli* distribution $X \in \{0, 1\}$ with

$$P(X = 1) = \theta \qquad P(X = 0) = 1 - \theta$$

then we have

$$\mathbf{E}[X] = \theta,$$

and

$$\mathbf{Var}\, X = \theta(1 - \theta).$$

Example 2 (Binomial)**:** The binomial distribution for $X \in \{0, 1, 2, \dots, n\}$, $X \sim Bin(\theta, n)$ arises from $Y_i \sim Bern(\theta)$ independent

$$X = \sum_{i=1}^{n} Y_i.$$

We must then have

$$P(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}.$$

One can prove this by counting the number of 1's or 0's in each n-string of experiments. We can thus see that

$$\mathbf{E}[X] = \mathbf{E}\left[\sum Y_i\right] = \sum \mathbf{E}[Y_i] = n\theta.$$

$$\mathbf{Var}(X) = \mathbf{Var}(\sum Y_i) = \sum \mathbf{Var}(Y_i) = n\theta(1 - \theta). \tag{1}$$

Example 3 (Geometric)**:** The geometric distribution $X \sim Geo(\theta)$ arises from given $Y_i \sim Bern(\theta)$ and independent with $X \in \{1, 2, \dots\}$, then

$$P(X = x) = \theta(1 - \theta)^{x-1} \tag{2}$$

and

$$\mathbf{E}[X] = \sum x\theta(1 - \theta)^{x-1} = \frac{1}{\theta}$$

Example 4 (Uniform)**:** The uniform distribution $X \sim Unif(\{1, 2, 3, \dots, n\})$ has $P(X = x) = n^{-1}$ and

$$\mathbf{E}[X] = \frac{n + 1}{2}$$

2 Aplications

2.1 Algorithm comlpexity

Suppose we have a program