Intro. Probability Theory

A Concise Review

Sample Space

- Sample space: The set of all possible outcomes of an experiment

  Flipping a coin: $S = \{H, T\}$

  Tossing a die: $S = \{1, 2, 3, 4, 5, 6\}$

  The lifetime of an elephant: $S = \{x \mid 0 \leq x < 100(\text{years?})\}$

- Any subset E of the sample space S is known as an event

  The event that a die lands on an even number

  $E = \{2, 4, 6\} \subset S$

Random Variables

- Real-valued functions defined on a sample space are known as random variables

  e.g., Let X denote the number of heads appearing when tossing 3 fair coins

  $P(X = 0) = P\{(T, T, T)\} = 1/8$

  $P(X = 1) = P\{(H, T, T), (T, H, T), (T, T, H)\} = 3/8$

  $P(X = 2) = P\{(H, H, T), (H, T, H), (T, H, H)\} = 3/8$

  $P(X = 3) = P\{(H, H, H)\} = 1/8$

- Discrete random variables: assume at most a countable number of possible values

  Continuous random variables whose set of possible values is uncountable

Distribution Functions

- The (cumulative) distribution function F of the random variable X is defined for all real numbers a, $-\infty < a < \infty$, by $F(a) = P(X \leq a)$

- F is nondecreasing; i.e., if $a < b$, then $F(a) \leq F(b)$

$$\lim_{a \to \infty} F(a) = 1$$

$$\lim_{a \to -\infty} F(a) = 0$$

$$P(a < X \leq b) = F(b) - F(a)$$

Discrete Random Variables: Examples

- Bernoulli random variable

  Consider a trial, whose outcome can be classified as either a "success" or a "failure". We define the random variable X that $X = 1$ corresponds to the outcome is success and $X = 0$, otherwise. Then X is said to be a Bernoulli random variable

$$P(X = 1) = p,\ P(X = 0) = 1 - p, P(X = 1) + P(X = 0) = 1$$

- Binomial random variable

  n independent trials are performed, each of which results in a "success" with probability p and a failure with probability $1 - p$. If X represents the number of successes of the n trials, then X is said to be a binomial random variable with parameters (n, p)

$$P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}, \quad i = 0, 1, \ldots, n$$

Discrete Random Variables: Examples

- Poisson random variable

  A random variable X, taking on one of the values $0, 1, 2, \ldots,$ is said to be a Poisson random variable with parameter $\lambda > 0$ if

$$P(X = i) = e^{-\lambda}\frac{\lambda^i}{i!}, \quad i = 0, 1, \ldots, n$$

- The Poisson random variable can be used as an approximation for a binomial random variable with parameters (n, p) when n is large and p is small enough so that np is of moderate size. (let $\lambda = np$)

$$\begin{aligned} P(X = i) &= \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{(n-i)!i!}\left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1)\cdots(n-i+1)}{n^i}\frac{\lambda^i}{i!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^i} \\ &\approx e^{-\lambda}\frac{\lambda^i}{i!} \quad \text{(for } n \text{ large and } \lambda \text{ moderate)} \end{aligned}$$

Discrete Random Variables: Examples

- Poisson random variable

  Hence, if n independent trials, each of which results in a "success" with probability p, are performed, then, when n is large and p is small enough to make np moderate, the number of successes occurring is approximately a Poisson random variable with parameter $\lambda = np$.

- Examples of random variables that usually obey the Poisson probability law
  - The number of misprints on a page of a book
  - The number of customers entering a post office on a given day

Continuous Random Variables

- X is a continuous random variable if there exists a non-negative function f, defined for $x \in (-\infty, \infty)$, having the property that for any set B of real number

probability density function (pdf)

$$P(x \in B) = \int_B f(x)\,dx$$

The pdf f must satisfy

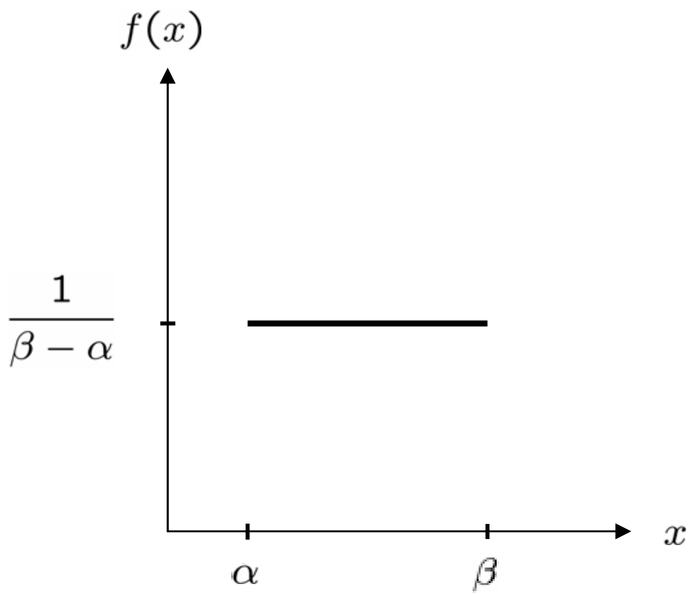
$$P\{x \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)\,dx = 1$$

Continuous Random Variables: Examples

- Uniform random variable

  X is a uniform random variable on the interval (α, β) if its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases}$$

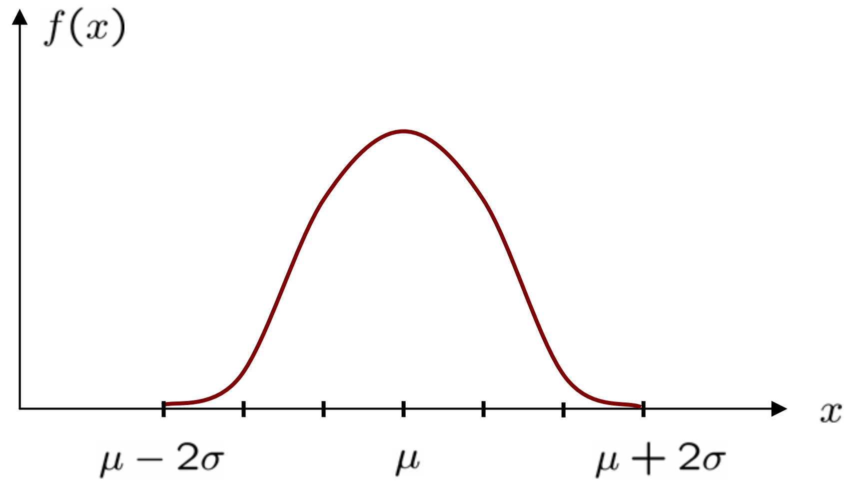


Continuous Random Variables: Examples

- Normal random variable

X is a normal random variable, with parameters μ and σ^2 if the density of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$



Joint Distribution Functions

- For random variables X and Y, the joint distribution function of X and Y is given by

$$F(a,b) = P(X \le a, Y \le b) \qquad -\infty < a, b < \infty$$

- When X and Y are both discrete random variables, it is convenient to define the joint probability mass function of X and Y by

$$p(x,y) = P(X = x, Y = y)$$

- Marginalization

$$p_X(x) = \sum_y p(x,y)$$

$$p_Y(y) = \sum_x p(x,y)$$

Joint Distribution Functions

- X and Y are said to be jointly continuous if there exists a function $f(x, y)$ defined for all x and y such that for every set C of pairs of real numbers

$$P\{(X,Y) \in C\} = \iint_{(x,y)\in C} f(x,y)dxdy$$

- Marginalization

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

Independent Random Variables

- X and Y are said to be independent if for any two sets of real numbers A and B

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

In the jointly continuous case the condition of independence is equivalent to

$$f(x, y) = f_X(x)f_Y(y) \qquad \text{for all } x, y$$

(Note that hereafter we will use p to represent the density function no matter that the random varibale(s) is discrete or continuous.)

Conditional Distributions

- The conditional probability mass/density function of X given $Y = y$ is given by

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}$$

For simplicity, we just write

$$p(x|y) = \frac{p(x,y)}{p(y)}$$

Analogously, we have

$$p(y|x) = \frac{p(x,y)}{p(x)}$$

$$\left.\begin{matrix} p(x|y) = \frac{p(x,y)}{p(y)} \\ p(y|x) = \frac{p(x,y)}{p(x)} \end{matrix}\right\} \quad p(y|x)p(x) = p(x|y)p(y)$$

Conditional Independence

- Consider, for example, three random variables X, Y, and Z. We say that X is conditionally independent of Y given $Z = z$ if

$$p(x|y,z) = p(x|z) \qquad (*)$$

On the other hand, from $(*)$, we have

$$\begin{aligned} p(x,y|z) &= p(x|y,z)p(y|z) \\ &= p(x|z)p(y|z) \end{aligned}$$

The conditional independence relation can be denoted by $X \perp\!\!\!\perp Y|Z$

Expectation

- The expectation of X is denoted by $E[X]$, and is defined by

$$E[X] = \sum_x xp(x) \qquad (X \text{ is discrete})$$

$$E[X] = \int_x xp(x)\,dx \qquad (X \text{ is continuous})$$

- Example: Calculate $E[X]$ of a Poisson random variable.

$$\begin{aligned} E[X] &= \sum_{i=0}^{\infty} ip(i) = \sum_{i=0}^{\infty} ie^{-\lambda}\frac{\lambda^i}{i!} = \sum_{i=1}^{\infty} ie^{-\lambda}\frac{\lambda^i}{i!} \\ &= \sum_{i=1}^{\infty} e^{-\lambda}\frac{\lambda^i}{(i-1)!} = \lambda e^{-\lambda}\sum_{i=1}^{\infty}\frac{\lambda^{i-1}}{(i-1)!} = \lambda \end{aligned}$$

Variance

- The variance of a random variable X, denoted by var(X), is defined by

$$\text{var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

- If a and b are constants, then

$$E[aX + b] = aE[X] + b$$

$$\text{var}(aX + b) = a^2\text{var}(X)$$

Covariance

- The covariance of any two random variables X and Y, denoted by $\mathrm{cov}(X,Y)$, is defined by

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

- Suppose X and Y are independent. Then

$$\begin{aligned}\mathrm{cov}(X,Y) &= E[XY]-E[X]E[Y] \\ &= \iint xyp(x,y)\,dxdy - \int xp(x)\,dx \int yp(y)\,dy \\ &= \iint xyp(x)p(y)\,dxdy - \int xp(x)\,dx \int yp(y)\,dy \\ &= 0\end{aligned}$$

- X, Y independent $\Rightarrow \mathrm{cov}(X,Y) = 0$

  $\mathrm{cov}(X,Y) = 0 \not\Rightarrow X$, Y independent

Correlation

- The correlation of two random variables X and Y, denoted by $\rho(X,Y)$, is defined by

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

  It can be shown that $-1 \le \rho(X,Y) \le 1$

- The correlation coefficeint is a measure of the degree of linearity between X and Y.

  $\rho(X,Y)$ near $+1$ or $-1 \Rightarrow$ high degree of linearity between X and Y

  $\rho(X,Y) > 0 \Rightarrow Y$ tends to increase when X does, and vice versa

  $\rho(X,Y) < 0 \Rightarrow Y$ tends to decrease when X does, and vice versa

  $\rho(X,Y) = 0 \Rightarrow X$ and Y are said to be uncorrelated

Central Limit Theorem

- Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed (i.i.d.) random variables each having mean μ and variance σ^2. Then the distribution of
$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$
tends to the standard normal as $n \to \infty$.

- Loosely put, the sum of a large number of independent random variables has a distribution that is approximately normal

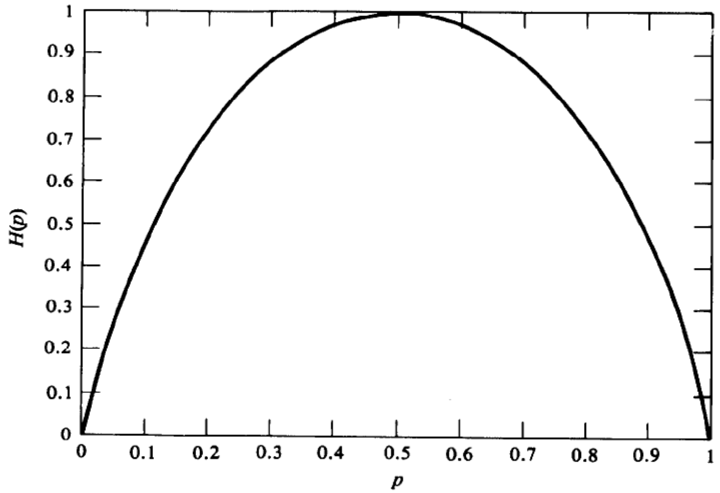
Entropy of a Random Variable

- Entropy (a measure of uncertainty of a random variable): The entropy $H(X)$ of a random variable X is defined by

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x),$$ where $\mathcal{X}$ is the alphabet of X.

- Example: Bernoulli random variable, $X = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } 1-p. \end{cases}$

$$H(X) = -p \times \log p - (1-p) \times \log(1-p) \equiv H(p)$$



Kullback-Leibler Distance and Mutual Information

- The Kullback-Leibler distance between two distributions is defined as (for simplicity, assume the discrete case)

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$$

- Consider two random variables X and Y. The mutual information $I(X,Y)$ is the relative entropy between the joint distribution and the product distribution, i.e,

$$I(X,Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$