Chapter 1: Probability

Dr.Ting Wei

Section 1.1: Concepts of probability

Question

- Can you think of an example of a random event?

Sample space

- A sample space is a list of outcomes of a random experiment, normally denoted by Ω.
- Examples:
  a) Tossing a coin: $\Omega = \{H, T\}$.

  b) Rolling a dice: $\Omega = \{1, 2, 3, 4, 5, 6\}$.

  c) Tossing two coins: $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$.

Question

- So the sample space is a set.
- For instance, when tossing a coin: $\Omega = \{H, T\}$.
- Lets consider the outcome 'head' that is the set $\{H\}$
- What is the relation between the two sets $\{H\}$ and $\{H, T\}$?

Basic rules of set operations

For any subsets $A, B \subseteq \Omega$,

- Commutative law: $A \cup B = B \cup A$, and $A \cap B = B \cap A$.
- Associative law: $(A \cup B) \cup C = A \cup (B \cup C)$, and $(A \cap B) \cap C = A \cap (B \cap C)$.
- Distributive law: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, and $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$.
- De Morgan's rules: $(A \cup B)^c = A^c \cap B^c$, and $(A \cap B)^c = A^c \cup B^c$.

Elementary results

- $A = B \Leftrightarrow A \subseteq B$ and $B \subseteq A$,
- $A \cap \emptyset = \emptyset$ and $A \cup \emptyset = A$,
- $A \cap \Omega = A$ and $A \cup \Omega = \Omega$,
- $A \cap A = A \cup A = A$,
- $A \cap B \subseteq A$ and $A \cap B \subseteq B$,
- $A \subseteq B \Rightarrow A \cap B = A$,
- $A \subseteq B^c \Rightarrow A \cap B = \emptyset$,
- $(A^c)^c = A$,
- $A \cup A^c = \Omega$, and $A \cap A^c = \emptyset$.

Events in a sample space

- An event is a subset of a sample space.

- Example: we toss a dice twice.
  $\Omega = \{(1,1),(1,2),(1,3)\ldots,(5,6),(6,6)\}$.

  The event 'the second toss is 6' is
  $A = \{(1,6),(2,6),(3,6),(4,6),(5,6),(6,6)\}$.

Question

- Suppose you are tossing a dice twice, so $\Omega = \{(1,1),(1,2),(1,3)\ldots,(5,6),(6,6)\}$.
- Can you spot a difference between the pair of events $(1,2)$, $(4,5)$ and $(1,2)$, $(1,3)$?

Disjoint events

- Events $A_1, \cdots A_n$ are disjoint if for all $i \neq j$, $A_i \cap B_j = \emptyset$.
- For instance if we toss a dice twice, the events $(1, 2)$, $(4, 5)$ are disjoint, but the events $(1, 2)$, $(1, 3)$ are not.

Sigma-algebras

- Intuition: given events A, B we want $A \bigcup B$ to be the event 'A or B'. We also want ' A does not happen' to be an event.

- A sigma-algebra on a sample space Ω is a collection $\mathcal{F}$ of subsets of Ω such that

  1) $\Omega \in \mathcal{F}$.

  2) if $A \in \mathcal{F}$, $A^c \in \mathcal{F}$.

  3) If $A_1, A_2, \ldots$ are in $\mathcal{F}$, $\displaystyle\bigcup_{i=1}^{\infty} A_i$ is in $\mathcal{F}$.

Example of a sigma-algebra

- $\{\emptyset, \Omega\}$, is a σ-algebra.
- $\{A, A^c, \emptyset, \Omega\}$ is a σ-algebra.
- The collection of all subsets of Ω is a σ-algebra and called the power set of Ω, which is denoted by $\mathbb{P}(\Omega)$.
- $\mathcal{F} = \mathbb{P}(\Omega)$ is the only σ-algebra we will consider from now on.

Probability spaces

- A probability function is a function $P : \mathcal{F} \to [0, 1]$. This maps each event to its probability of occurrence.

- We also impose the probability axioms:

  1) $P(\emptyset) = 0$.

  2) $P(\Omega) = 1$.

  3) (Countable additivity)

     If $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \ .$$

- The triple $(\Omega, \mathcal{F}, P)$ is called a probability space.

Examples of probability spaces

- Fair coin toss $\Omega = \{H, T\}$, $\mathcal{F} = \mathbb{P}(\Omega)$, $P(H) = P(T) = 1/2$, $P(\emptyset) = 0$, $P(\Omega) = 1$.

- Unfair coin toss: choose $0 \leq p \leq 1$ and put $\Omega\{(H, T)\}$, $\mathcal{F} = \mathbb{P}(\Omega)$, and define P by $P(\emptyset) = 0$, $P(\Omega) = 1$ and

$$P(\{H\}) = p, \; P(\{T\}) = 1 - p \; .$$

- 6-sided fair dice: $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \mathbb{P}(\Omega)$, $P(A) = |A|/6$, where $|A|$ is the size of the subset A.

Basic results on probability

- One can prove the following properties

  Let P be a probability function on a sigma-algebra $\mathcal{F}$ and let $A, B \in \mathcal{F}$. Then:

  i) $P(A^c) = 1 - P(A)$

  ii) If $A \subseteq B$ then $P(A) \leq P(B)$

  iii) $P(B \setminus A) = P(B) - P(B \cap A)$

  iv) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Independence

- Idea: two events are independent when there is no connection between them.

- For instance, if we toss a coin twice, it is natural to think that the result of the first toss is independent from, or does not affect, the result of the second.

- Two events A and B are called independent if

$$P(A \cap B) = P(A)P(B).$$

Example

- Consider the example of rolling a single 6-sided fair dice. Thus $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \mathbb{P}(\Omega)$ and $P(\{\omega\}) = 1/6$ for all $\omega \in \Omega$.

- Let $E = \{2, 4, 6\}$ be the event that the roll is even and $T = \{2\}$ be the event that the roll is 2. Is E independent of T?

Mutual independence

- A collection of events are mutually independent if and only if the probability of any intersection is equal to the product of the probabilities of the events included in the intersection.
- Suppose that for each i in some set I we have an event A_i. The family of events $\{A_i : i \in I\}$ is called independent if for any n and any *distinct* elements $i_1, \ldots, i_n$ of I we have

$$P(A_{i_1} \cap \cdots \cap A_{i_n}) = \prod_{j=1}^{n} P(A_{i_j}).$$

Conditional probability: motivation

- Motivation: we want to measure the probability of an event A occurring given that another event B occurs.
- Example: In a customer survey we found that 60% of the customers used the brand Favorite (F), while 30% used the brand Super (S). 15% of the customers used both brands.
- How many percent of the Favorite users use Super, and how many percent of the Super users use Favorite?

Conditional probability: definition

- Idea: change the sample space to B and scale all the probabilities accordingly.

- Let A and B be events with $P(B) \neq 0$. The conditional probability of A given B, written $P(A|B)$, is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Back to the example

- In a customer survey we found that 60% of the customers used the brand Favorite (F), while 30% used the brand Super (S). 15% of the customers used both brands. How many percent of the Favorite users use Super, and how many percent of the Super users use Favorite?

- We want to find $P(S|F)$ and $P(F|S)$.

Bayes' formula

Figure: Thomas Bayes (c. 1701 – 7 April 1761)

- Let A and B be events with nonzero probability. Then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Law of total probability

- Suppose $\mathcal{E} = \{E_1, \ldots, E_n\}$ is a partition of the sample space Ω and A is any event. Then,
- $P(A) = \sum_{i=1}^{n} P(A \cap E_i)$.
- $P(A) = \sum_{i=1}^{n} P(A|E_i)P(E_i)$, where the sum is over all i such that $P(E_i) \neq 0$.

Bayes' Theorem

Theorem

Let $\mathcal{E} = \{E_1, \ldots, E_n\}$ *be a partition of* Ω*, and* A *be an event. Then*

$$P(E_j|A) = \frac{P(E_j)P(A|E_j)}{\sum_{i=1}^{n} P(E_i)P(A|E_i)}$$

where $P(E_j) > 0$ *for all* $1 \leq j \leq n$ *and* $P(A) > 0$*.*

Example

- A computer centre has three printers: A, B, and C.
- Documents are routed to printers A, B, and C with probability 0.5, 0.3, and 0.2 respectively.
- Printers A, B, and C, jam with probability 0.04, 0.05, and 0.03, respectively.
- Now you find your program crashes because of printer jamming. What is the probability that printer A is the culprit?

Independence and conditional probability

- If $P(B) \neq 0$, then A and B are independent if and only if $P(A|B) = P(A)$.

- Proof: If A and B are independent,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

If $P(A|B) = P(A)$ then

$$P(A \cap B) = P(A|B)P(B) = P(A)P(B)$$

so by definition A and B are independent.

Summary

- The sample space Ω is the set of all outcomes, and an event is a subset of the sample space.
- A probability space is the triple $(\Omega, \mathcal{F}, P)$ where the probability function $P : \mathcal{F} \to [0, 1]$ satisfies certain axioms.
- Two events A and B are independent if $P(A \cap B) = P(A)P(B)$.
- Mutual independence of collection of events.
- The conditional probability of A given B is $P(A|B) = \frac{P(A \cap B)}{P(B)}$.
- Bayes' rule: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.
- Law of total probability.

Section 1.2: Random variables, Part I

Question

- Can you make up an example of a random variable?

The idea of random variable

- Motivation: we wish to study the numbers which come as a result of random outcomes.

- Example : Let Ω be the sample space of outcomes from two rolls of a dice. Define a function $X(\omega_1, \omega_2) = \omega_1 + \omega_2$ where ω_1 is the result of the first roll and ω_2 is the result of the second roll. Any outcome of the two rolls leads to a real number.

- This example gives a function X on the sample space Ω leading to a real number. We call any such function a random variable.

Definition of random variable

- A random variable X is a function defined on the sample space Ω which returns a well defined real number from any outcome.
- The rigorous definition requires that the above function is 'measurable'. All the functions in this course will be, so we do not elaborate further.

Distribution of a random variable: motivation

- Consider a random variable $X : \Omega \to \mathbb{R}$. Say we want to know the probability that X takes a value less than x. In term of events we have

$$\{\omega : X(\omega) \leq x\} \Leftrightarrow \omega \in X^{-1}((-\infty, x])$$

  where X^{-1} is the inverse of the function X.

- The probability of this event is denoted more briefly by $P(X \leq x)$.

Distribution of a random variable

- Let X be a random variable. The cumulative distribution function (CDF) or distribution function of X is the function $F_X : \mathbb{R} \to \mathbb{R}$ defined by

$$F_X(x) = P(X \leq x).$$

PMF of discrete random variables

- Let X be a random variable. We say X is discrete if it is defined on a discrete sample space Ω with values in $\mathbb{R}$.

- Let X be a discrete random variable. The probability mass function or PMF or probability function of X is the function $p_X(x_i) = P(X = x_i)$, $i = 1, 2, \ldots, n$.

- Obviously, $p_X(x_i) \geqslant 0$ and $\sum_{i=1}^{n} p_X(x_i) = 1$.

Infinite but discrete

- A discrete random variable can be infinite but discrete.
- Example: A coin is tossed until a head occurs. Can you draw the sample space and define a random variable on that?

Discussion questions

- Suppose that $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and

  $P(X = \omega_1) =$0.2, $P(X = \omega_2) =$0.1, $P(X = \omega_3) =$0.7.

  $X(\omega_1) = 1$, $X(\omega_2) = 2$, $X(\omega_3) = 3$.

- Can you draw the CDF ?

- Recall the CDF is $F_X(x) = P(X \leq x)$ and the PMF is $p_X(x_i) = P(X = x_i)$, $i = 1, 2, \ldots, n$.

- What is the relation between CDF and PMF for a discrete random variable X taking values x_i, $i \in \mathbb{N}$?

Expectation of a discrete random variable

- Idea: the expected value of a discrete random variable X is a weighted average of the values that X takes, each value is weighted according to its probability.

- Let X be a discrete random variable with a finite list $\{x_1, ..., x_n\}$ of possible values. The expectation of X is defined by

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i).$$

Discrete uniform distribution

- A discrete random variable X with n possible outcomes $x_1, \ldots, x_n$ is said to have a discrete uniform distribution if the PMF of X is given by

$$P(X = x_i) = \frac{1}{n} \qquad i = 1, \ldots, n\,.$$

- Example: if we roll a fair dice, the outcomes $1, \ldots, 6$ have equal probability of occurring, hence the random variable X 'number of dots observed on the upper face of the dice' has a uniform discrete distribution with PMF $P(X = i) = \dfrac{1}{6}$ $i = 1, \ldots, 6$.
- $E(X) = \frac{1}{n} \sum_{i=1}^{n} x_i$.

Bernoulli random variables

- The Bernoulli distribution is used to model an experiment with only two possible outcomes, often referred to as 'success' and 'failure', or 0 and 1.
- A discrete random variable X has a Bernoulli distribution with parameter $0 \leq p \leq 1$ if its probability mass function is

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

- We write $X \sim \text{Ber}(p)$ if X is a Bernoulli random variable with parameter p, and $E(X) = p$.

Binomial random variables

- A random variable with PMF

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

  $k = 0, 1, \ldots, n$ is called a binomial random variable with parameters n and p.

- We write $X \sim \mathrm{Bin}(n, p)$ if X is a binomial random variable with parameters n and p, and $E(X) = np$.

Example

- Suppose we have a coin which has probability p of showing heads. We make n independent tosses of the coin, recording the result, and count the number of heads obtained in the n tosses.

- This is modelled by a binomial random variable with parameters n and p.

Poisson random variable

- Let $\lambda > 0$. A random variable with PMF
$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$
for $k = 0, 1, 2, \ldots$ is called a Poisson random variable with parameter λ.

- We write $X \sim \text{Pois}(\lambda)$ if X is a Poisson random variable with parameter λ, and $E(X) = \lambda$.

- Under certain independence assumptions it can be shown that the number of events happening in time t is a Poisson random variable, e.g. for the number of road accidents, industrial accidents in a given unit of time. Also, the number of calls received by a switchboard.

Example

- The number of industrial accidents at a particular manufacturing plant is found to average three per month, and it is likely to follow a Poisson distribution.

Geometric random variables

- A random variable with PMF

$$P(X = k) = (1-p)^{k-1}p$$

  $k = 1, \ldots, n$ is called a geometric random variable with parameters p.

- We write $X \sim \text{Geom}(p)$ if X is a geometric random variable with parameters p, and $E(X) = \frac{1}{p}$.

- Recall the example: A coin is tossed until a head occurs.

Summary

- Notion of random variable $X : \Omega \to \mathbb{R}$.
- Cumulative distribution function (CDF) $F_X(x) = P(X \leq x)$.
- Probability mass function (PMF) of a discrete random variable.
- Examples of discrete random variables: Uniform, Bernoulli, Binomial, Poisson, Geometric.

Section 1.3: Random variables, Part II

PDF of continuous random variables

- A continuous random variable is a random variable X such that the CDF $F_X(x)$ satisfies

$$F_X(x) = P(X \leqslant x) = \int_{-\infty}^{x} f_X(y)dy$$

for some function $f_X : \mathbb{R} \to \mathbb{R}$.

- This function $f_X(x)$ is a non-negative integrable function called the probability density function or density function or PDF of X.

- By fundamental theorem of calculus: $\dfrac{dF_X(x)}{dx} = f_X(x)$.

- For all $x \in \mathbb{R}$, $f_X(x) \geqslant 0$.

PDF of continuous random variables, cont.

- It follows that if X is a continuous random variable with PDF $f_X(x)$ we have

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$$

for any a, b with $a \leq b$.

- and

$$\int_{-\infty}^{\infty} f_X(x)dx = 1.$$

True or False?

- From previous slide:

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

- Is it true or false that $P(X = a) = f_X(a)$?

Expectation of a continuous random variable

- Let X be a continuous random variable with PDF $f(x)$.
- We define the expected value $E(X)$ as

$$E(X) = \int_{-\infty}^{\infty} x\, f(x) dx \ .$$

Continuous uniform random variable

- If X is a continuous variable with values in the interval $[a, b]$, with density function

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b, \\ 0, & \text{otherwise} \end{cases}$$

  Then X is called a continuous uniform random variable on the interval $[a, b]$.

- We write $X \sim \mathrm{U}(a, b)$ if X is a continuous uniform random variable with parameters a and b, and $E[X] = \frac{a+b}{2}$.

- The cumulative distribution function is

$$F_X(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

Example of continuous uniform random variable

- Suppose a train arrives at a subway station regularly every 10 minutes.
- If a passenger arrives at the station without knowing the timetable, then the waiting time to catch the train is uniformly distributed.

Exponential random variable

- The positive random variable X with density

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

  for some positive parameter λ is called an exponential random variable.

- We write $X \sim \mathrm{Exp}(\lambda)$ if X is an exponential random variable with parameters λ, and $E[X] = \frac{1}{\lambda}$.

- The cumulative distribution function is

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Example of exponential random variable

- The lifetime of a certain type of car batteries follows an exponential distribution. It is known that the average lifetime is 5 years. What's the probability that a particular battery of this type lasts for more than 7 years?

Normal (Gaussian) random variables

- A continuous random variable X is said to have a normal (Gaussian) distribution if its density is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

for some μ and positive $\sigma > 0$.

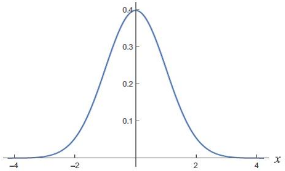
- We write $X \sim N(\mu, \sigma^2)$ if X is a normal random variable with parameters μ and σ, and $E[X] = \mu$.

Normal distribution

- The standard normal random variable has PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

The graph of PDF is shown here for $\mu = 0$ and $\sigma = 1$ (in this case is $N(0, 1)$).



Normal, cont.

- Thus

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

The probability that X is between a and b is hence equal to the area under the density function between a and b.

- It can be shown that

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$

Normal, cont.

- The CDF of the normal random variable X is

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

  There is no explicit formula to solve this integral, it can only be solved by numerical methods.

Find probabilities of standard normal

- We usually use $\Phi(x)$ to denote the CDF of a standard normal random variable,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy$$

- It can be shown that $\Phi(1.96) \approx 0.975$. Thus if $X \sim N(0,1)$, $P(X \leq 1.96) \approx 0.975$.
- $P(a \leq X \leq b) = \Phi(b) - \Phi(a)$.
- $\Phi(x) = 1 - \Phi(-x)$.
- Find $\Phi(x)$ using standard normal table or programming.

```
prob <- pnorm(x,mu,sd)
```

Question

- Let $X \sim N(\mu, \sigma^2)$. How to find $P(X \leqslant x)$ from the statistical table?

Answer

- If $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$.
- **Example**: Assume $X \sim N(10, 4)$, find $P(8 < X < 11)$.

Variance

- The variance $\mathrm{var}(X)$ of a random variable X is defined by

$$\mathrm{var}(X) = E\left[(X - E[X])^2\right]$$

- The variance is a measure of the deviation from the expected value.

- The standard deviation $\sigma(X)$ of a random variable X is defined as $\sigma(X) = \sqrt{\mathrm{var}(X)}$.

Properties of expectation and variance

Let X, Y be random variables and $a, b, c \in \mathbb{R}$. Then

- $E(aX + bY + c) = aE(X) + bE(Y) + c$.
- $\text{var}(aX + bY + c) = a^2 \text{var}(X) + 2ab \text{ cov}(X, Y) + b^2 \text{var}(Y)$ where $\text{cov}(X, Y)$ is the covariance between X and Y .
- We will return to the covariance $\text{cov}(X, Y)$ later. For now just note that the variance is not linear.
- $\text{var}(X) = E(X^2) - E(X)^2$.

Summaries of expectation and variance

- If X follows a discrete uniform distribution, $E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\mathrm{var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - E(X))^2$.
- $X \sim \mathrm{Ber}(p)$, $E(X) = p$ and $\mathrm{var}(X) = p(1-p)$.
- $X \sim \mathrm{Bin}(n,p)$, $E(X) = np$ and $\mathrm{var}(X) = np(1-p)$
- $X \sim \mathrm{Pois}(\lambda)$, $E(X) = \lambda$ and $\mathrm{var}(X) = \lambda$
- $X \sim \mathrm{Geom}(p)$, $E(X) = \frac{1}{p}$ and $\mathrm{var}(X) = \frac{1-p}{p^2}$

Summaries of expectation and variance

- $X \sim \mathrm{U}(a, b)$, $E(X) = \frac{a+b}{2}$ and $\mathrm{var}(X) = \frac{(b-a)^2}{12}$
- $X \sim \mathrm{Exp}(\lambda)$, $E(X) = \frac{1}{\lambda}$ and $\mathrm{var}(X) = \frac{1}{\lambda^2}$
- $X \sim N(\mu, \sigma^2)$, $E(X) = \mu$ and $\mathrm{var}(X) = \sigma^2$

Functions of random variables

- Motivating question: given a (reasonably nice) function f and a random variable X whose distribution is known, what is the distribution of $f(X)$?

- Suppose g is differentiable and strictly increasing or strictly decreasing and X is a continuous random variable with density function f_X. Then the density function of $Y = g(X)$ is

$$f_Y(y) = \left| \frac{dg^{-1}}{dy} \right| f_X(g^{-1}(y))$$

Expectation of functions of random variables

- Let X be a discrete random variable and g be a function such that $E[g(X)]$ exists. Then

$$E[g(X)] = \sum_i g(x_i) P(X = x_i) \ .$$

- Let X be a continuous random variable with density function f_X. Suppose g is a function and $E[X]$, $E[g(X)]$ exist. Then

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \ .$$

Example

- A process for refining sugar yields up to 1 ton of pure sugar per day, but the actual amount produced X is a random variable because of machine breakdowns. Suppose that X has density function given by

$$f(x) = \begin{cases} 2x, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

- The company is paid at the rate of £ 300 per ton for the refined sugar, but it has also has a fixed overhead cost of £ 100 per day. Thus the daily profit, in hundreds of pounds, is $Y = 3X - 1$.

- Find the probability density function of Y.

Summary

- Probability density function (PDF) of continuous random variables.
- Examples of continuous random variables: Uniform, Exponential, Normal.
- Expectation and variance.
- Functions of random variables.

Section 1.4: Independence and joint distributions

Jointly distributed random variables

- Idea: when we study two random variables X and Y, we first make a list of all possible values X and Y can have. Then we consider all the possible combinations that can occur, and figure out the probability of such combinations.

- Two random variables on the same sample space are called jointly distributed.

Joint cumulative distribution function

- Let X and Y be jointly distributed random variables. The joint CDF $F_{X,Y}$ of X and Y is the function

  $$F_{X,Y}(x, y) = P(X \le x \text{ and } Y \le y).$$

- This is the probability of the event

  $$X^{-1}\left((-\infty, x]\right) \cap Y^{-1}\left((-\infty, y]\right) .$$

Independent random variables

- Imposing the condition that these two events are independent we find

$$F_{X,Y}(x,y) = P(X^{-1}(-\infty,x] \cap Y^{-1}(-\infty,y]) =$$
$$= P(X^{-1}(-\infty,x])P(Y^{-1}(-\infty,y]) = F_X(x)F_Y(y) \; .$$

- In this case, the jointly distributed random variables X and Y are called independent

Overview

Given a random variables X and Y,

CDF $P(X \leq x)$	
DISCRETE	**CONTINOUS**
PMF $P(X = x)$	Density $f(x)$

Joint CDF $P(X \leq x$ and $Y \leq y)$	
DISCRETE	**CONTINOUS**
Joint PMF	Density $f(x, y)$

Can you guess what goes in place of the dots ?

Jointly distributed discrete random variables

- Discrete random variables on the same sample space are called jointly discrete if they are both discrete.
- Let X and Y be jointly distributed discrete random variables. The joint probability mass function or joint PMF of X and Y is

$$p_{X,Y}(x,y) = P(X = x \text{ and } Y = y).$$

Example

- In a town there are two different firms with about the same number of customers. Let X denote which firm is chosen by a randomly selected customer, and Y the number of days it takes to process an order.
- The joint PMF is

	$Y = 1$	$Y = 2$	$Y = 3$
$X = 1$	0.1	0.1	0.3
$X = 2$	0.2	0.1	0.2

Marginal distributions: discrete case

- Let X, Y be jointly distributed discrete random variables with joint PMF $p_{X,Y}$.
- The marginal probability mass function of X is

$$p_X(x_i) = P(X = x_i) = \sum_j P(X = x_i \text{ and } Y = y_j).$$

- Fact: X and Y are independent if and only if $p_{X,Y}(x, y) = p_x(x)p_Y(y)$.

Example and Question

- In the previous example the marginals are

	$Y = 1$	$Y = 2$	$Y = 3$	$P(X = x)$
$X = 1$	0.1	0.1	0.3	0.5
$X = 2$	0.2	0.1	0.2	0.5
$P(Y = y)$	0.3	0.2	0.5	

Question: are X and Y independent?

Jointly distributed continuous random variables

- Random variables X and Y on the same sample space are called jointly continuous if there is a function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ such that

  $$F_{X,Y}(u,v) = \int_{-\infty}^{u} \int_{-\infty}^{v} f_{X,Y}(x,y)dxdy.$$

  In this case we call $f_{X,Y}$ the joint density function or joint PDF of X and Y.

- We can recover the joint PDF from the joint CDF by differentiation: at points at which $f_{X,Y}$ is continuous we have

  $$f_{X,Y} = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}.$$

Marginal distributions: continuous case

- Let X, Y be jointly continuous random variables with joint density function $f_{X,Y}$.
- The marginal density function of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy.$$

- Fact: jointly continuous random variables are independent if and only if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Example. The bivariate normal distribution

- Two random variables X and Y are said to have a bivariate normal distribution if their joint probability density function is given by

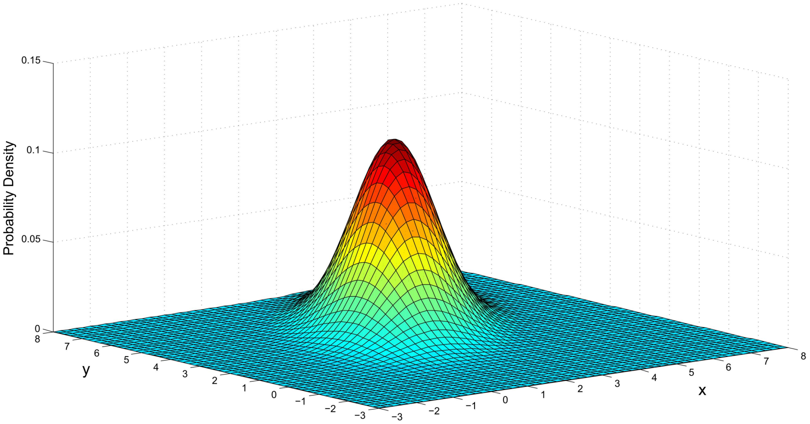
$$f_{X,Y}(x,y) =$$

$$= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \right.\right.$$
$$\left.\left. + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\}\right]$$

for $-\infty < x, y < \infty$, where ρ is the correlation between X and Y.

- There are five parameters: μ_x, μ_y, σ_x, σ_y, ρ.

The bivariate normal distribution



A vertical cross-section through the surface at $x = x_0$ or at $y = y_0$ is a normal density curve.

Example

- Let X, Y have the bivariate normal distribution with joint PDF as on slide of previous example

$$f_{X,Y}(x,y) =$$

$$= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \right.\right.$$

$$\left.\left. + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\}\right]$$

- The marginal distributions of X and Y are normal with means μ_X, μ_Y and variances σ_X^2 and σ_Y^2.

- If $\rho = 0$ then $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, so X and Y are independent.

Conditional distributions: discrete case

Let X, Y be discrete random variables.

- The conditional probability mass function $p_{X|Y}$ is

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x \text{ and } Y = y)}{P(Y = y)}.$$

  This function only makes sense for y such that $P(Y = y) \neq 0$, since otherwise the conditional probability isn't defined.

- The conditional expectation of X given Y is

$$E(X|Y = y) = \sum_i x_i p_{X|Y}(x_i|y).$$

Example

- In the previous example:

	$Y = 1$	$Y = 2$	$Y = 3$	$P(X = x)$
$X = 1$	0.1	0.1	0.3	0.5
$X = 2$	0.2	0.1	0.2	0.5
$P(Y = y)$	0.3	0.2	0.5	

- Find $p_{X|Y}(1|1)$, $p_{X|Y}(1|2)$, $p_{X|Y}(1|3)$, $p_{X|Y}(2|1)$, $p_{X|Y}(2|2)$, $p_{X|Y}(2|3)$.

Conditional distributions: continuous case

Let X, Y be continuous random variables.

- The conditional density function is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

- The conditional expectation of X given Y is

$$E(X|Y=y) = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)dx.$$

Summary

- Jointly distributed random variables and their joint CDF.
- Independent random variables.
- Joint PMF of jointly discrete random variables,
- Joint PDF of jointly continuous random variables.
- Marginal distributions.
- Conditional distributions and expectations.

Section 1.5: Random sampling

Question

- We obtain a dataset of ten elements by tossing a fair coin ten times and recording the result of each toss.
- Can you model this experiment by a sequence of random variables?
- If so, what are they and which characteristics do they have?

Random sample: the idea

- Often the elements of a dataset are repeated measurements of the same quantity.
- We interpret the outcome of an experiment as a realization of some random variables.
- If the measurements are obtained under the same experimental conditions, it is justified to assume that the underlying probability distributions are the same.
- If the successive measurements do not influence each others, we can assume that the random variables are mutually independent.

Random sample definition

- A random sample is a collection of random variables $X_1, \ldots, X_n$ that have identical probability distributions and are mutually independent.

Sample statistic: the idea

- Given a dataset $x_1, \ldots, x_n$, many empirical summaries can be written, such as the mean $(x_1 + \cdots + x_n)/n$.
- Such summaries are functions $h(x_1, \ldots, x_n)$.
- Since datasets are modelled as realization of a random sample $X_1, \ldots, X_n$, an object $h(x_1, \ldots, x_n)$ is a realization of the corresponding random variable $h(X_1, \ldots, X_n)$.

Example

- Consider the random sample associated to rolling a fair dice 3 times.
- Suppose in an experiment we obtain the dataset $\{2, 6, 4\}$.
- We are interested in the mean of the first two rolls.
- What is the numerical summary and the corresponding statistics?

Sample statistic: definition

- A statistic is a function of the sample random variables.
- Thus a statistic is itself a random variable. The distribution of a statistic is called its sampling distribution.
- If a statistical model adequately describes the dataset at hand, then the sample statistics corresponding to the empirical summaries should reflect the corresponding features of the model distribution.

Example

1. The sample mean is $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$.

2. The sample variance is $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Sample mean from normal population

- Let $X_1, \ldots, X_n$ be a random sample from a normal population $N(\mu, \sigma^2)$.

- The expectation of the sample mean is

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

- Since X_i are independent,

$$\mathrm{var}(\bar{X}) = \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Sample variance from a normal random sample

- Recall the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
- It can be proved that the following important relationship holds:

$$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}.$$

- This can be considered as an example of a χ^2 random variable.

Connecting the t, F and normal distributions

- Let $X_1, \ldots, X_n$ be n independent and identically normally distributed variables with distribution $N(\mu_X, \sigma_X^2)$.
- Let $Y_1, \ldots, Y_m$ be m independent and identically normally distributed variables with distribution $N(\mu_Y, \sigma_Y^2)$. Suppose each X_i is independent of each Y_j.
- Then the following relationships hold
  - $\bar{X} \sim N(\mu_X, \sigma_X^2/n)$.
  - $\dfrac{\bar{X} - \mu_X}{S_X/\sqrt{n}} \sim t_{n-1}$.
  - $\dfrac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{n-1,m-1}$.

Convergence in distribution

- Suppose we have a sequence of random variables $X_1, X_2, \ldots$ with CDFs F_{X_1}, F_{X_1}, and X is a random variable with CDF F_X.

- We say the X_i converge in distribution to X as $n \to \infty$ and write

$$X_n \xrightarrow{D} X \quad \text{as} \;\; n \to \infty$$

if $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$ for all $x \in \mathbb{R}$.

Central limit theorem (CLT)

- Central Limit Theorem (CLT): Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with mean μ and variance σ^2. Let $\bar{X}_n$ be the sample mean of $X_1, \ldots, X_n$. Then

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1) \quad \text{as} \quad n \to \infty \, .$$

- This is useful because of its generality: it implies that the sample mean of an independent sample from any distribution with finite mean and variance has approximately the $N(\mu, \sigma^2/n)$ distribution for large n.

Central limit theorem (CLT), cont.

- Let $S_n = X_1 + \cdots + X_n$, so $\bar{X}_n = S_n/n$.
- Thus $E(S_n) = n\mu, \quad \mathrm{var}(S_n) = n\sigma^2$.
- We note that

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{S_n/n - \mu}{\sigma/\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n - E(S_n)}{\sqrt{\mathrm{var}(S_n)}}.$$

- Thus by the CLT

$$\frac{S_n - E(S_n)}{\sqrt{\mathrm{var}(S_n)}} \xrightarrow{D} N(0,1).$$

Example

- Assume that 400 customers come to a shop every day and that their purchases are independent. They often buy milk, and we assume that the number of liters bought by each customer is a random variable X with distribution

$$P(X = 0) = 0.3 \quad , P(X = 1) = 0.5, \quad P(X = 2) = 0.2 \, .$$

- Find an approximate value for the probability that the shop sells between 341 and 390 liters.

Example: solution (fill gaps)

- We call the total number of liters sold $S =$________________
  We need to compute
  $$P(341 \leq S \leq 390) = P(\_\_\_\_\_\_) - P(\_\_\_\_\_\_)\ .$$
- We can find approximate values for these probabilities from the central limit theorem, but we need to know $E(S)$ and $\mathrm{var}(S)$.
- We first compute $E(X)$ and $\mathrm{var}(X)$:
  $$\begin{aligned} E(X) &= \underline{\hspace{5cm}} \\ E(X^2) &= \underline{\hspace{5cm}} \\ \mathrm{var}(X) &= \underline{\hspace{5cm}} \end{aligned}$$
- Therefore
  $$E(S) = \underline{\hspace{7cm}}$$

Example: solution, cont.

- Since $X_1, \ldots, X_{400}$ are independent we have

  $$\mathrm{var}(S) = \underline{\hspace{6cm}} \qquad \text{hence} \quad \sigma = \underline{\hspace{2cm}}\,.$$

- As $n = 400$ is large, by the CLT,

  $$\frac{S - E(S)}{\sqrt{\mathrm{var}(S)}} \sim N(0,1).$$

- We calculate

  $$P(S \leq 390) - P(S \leq 340) \approx \Phi\left(\underline{\hspace{2cm}}\right) - \Phi\left(\underline{\hspace{2cm}}\right) =$$
  $$= \underline{\hspace{7cm}}$$

  The probability that the customers buy between 341 and 390 liters of milk is hence about 91%.

Normal approximations

- Let $B \sim \text{Bin}(n, p)$. Then

$$\frac{B - np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0, 1)$$

as $n \to \infty$.

- Note that a binomial distribution can always be viewed as a sum of n independent Bernoulli distributions, and the central limit theorem can be applied directly to the distribution.

Continuity correction

- Continuity correction should be applied for normal approximations of binomial probabilities.
- Let $B \sim \text{Bin}(n, p)$ and $Y \sim N(np, np(1-p))$, continuity correction states
  - $P(B \leq x) \approx P(Y \leq x + 0.5)$
  - $P(B < x) \approx P(Y \leq x - 0.5)$
  - $P(B > x) \approx P(Y \geq x + 0.5)$
  - $P(B \geq x) \approx P(Y \geq x - 0.5)$
- For n sufficiently large, the following result holds

$$P(B \leq x) \approx \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

Example (fill gaps)

- A fair coin is tossed 1000 times. Find the probability P that heads occur exactly 531 times.

- Let B be the random variable counting the number of heads, so $B \sim \text{Bin}(1000, 1/2)$. We then have an exact expression for this probability:

$$P(B = 531) = \underline{\qquad\qquad}$$

- Unfortunately $\binom{1000}{531}$ is nearly 10^{300}, too large to deal with directly on an ordinary calculator. Instead (from the context of the rest of the question) we should try to use a normal approximation.

Example, cont.

- The normal approximation says that $P(B \leq x) \approx P(Y \leq x + 0.5)$ where $Y \sim N(500, 250)$. Thus the probability we want is

$$\begin{aligned} P(B = 531) &= P(B \leq 531) - P(B \leq 530) \\ &\approx P(\underline{\qquad\qquad}) - P(\underline{\qquad\qquad}) \\ &= \underline{\qquad\qquad\qquad\qquad\qquad\qquad} \end{aligned}$$

Summary

- Random sample.
- Statistic and sampling distributions.
- Samples from a normal population.
- Connecting the t, F and normal distributions.
- Central limit theorem and normal approximations.