202A Notes

Chapter 1 Probability Theory

1.1 Set Theory

Def 1.1.1 (Sample space)
The set, S, of all possible outcome of a particular experiment is called the sample space for the experiment.

Def 1.1.2 (Event)
An event is any collection of possible outcomes of an experiment, that is, any subset of S (including S itself).

Th 1.1.4 (Properties of sets)
For any three events, A, B, and C, define on a sample space S,

1. Commutativity:
   $A \cup B = B \cup A$; $A \cap B = B \cap A$
2. Associativity:
   $A \cup (B \cup C) = (A \cup B) \cup C$; $A \cap (B \cap C) = (A \cap B) \cap C$
3. Distributive Laws:
   $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$; $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
4. DeMorgan's Laws:
   $(A \cup B)^c = A^c \cap B^c$; $(A \cap B)^c = A^c \cup B^c$

Def 1.1.5 (Disjoint sets)
Two events A and B are disjoint (or mutually exclusive) if $A \cap B = \emptyset$. The events $A_1, A_2, ...$ are pairwise disjoint (or mutually exclusive) if $A_i \cap A_j = \emptyset$ for all $i \neq j$.

Def 1.1.6 (Partition)
If $A_1, A_2, ...$ are pairwise disjoint and $\cup_{i=1}^{\infty} = S$, then the collection $A_1, A_2, ...$ forms a partition of S.

1.2 Basics of Probablility Theory

Def 1.2.1 (σ-Algebra)
A collection of subsets of S is called a sigma algebra (or Borel field), denoted by $\mathcal{B}$, if it satisfies the following three properties:

1. $\emptyset \in \mathcal{B}$
2. If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$ ($\mathcal{B}$ is closed under complementation)

3. If $A_1, A_2, \dots \in \mathcal{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$ ($\mathcal{B}$ is closed under countable unions)

Def 1.2.4 (Probability function)
Given a sample space S and an associated sigma algebra $\mathcal{B}$, a probability function is a function P with domain $\mathcal{B}$ that satisfies:

1. $P(A) \geq 0$ for all $A \in \mathcal{B}$
2. $P(S) = 1$
3. If $A_1, A_2, \dots \in \mathcal{B}$ are pairwise disjoint, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

Th 1.2.6 (Simple definition of probability function)
Let $S = \{s_1, \dots, s_n\}$ be a finite set. Let $\mathcal{B}$ be any sigma algebra of subsets of S. Let $p_1, \dots, p_n$ be nonnegative numbers that sum to 1. For any $A \in \mathcal{B}$, define $P(A)$ by

$$P(A) = \sum_{i:s_i \in A} p_i.$$

(The sum over an empty set is defined to be 0.) Then P is a probability function on $\mathcal{B}$. This remains true if $S = \{s_1, s_2, \dots\}$ is a countable set.

Th 1.2.8 (Properties of the probability funciton I)

1. $P(\emptyset) = 0$
2. $P(A) \leq 1$
3. $P(A^c) = 1 - P(A)$

Th 1.2.9 (Properties of the probability funciton II)
If P is a probability function and A and B are any sets in $\mathcal{B}$, then

1. $P(B \cap A^c) = P(B) - P(A \cap B)$
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3. If $A \subset B$, then $P(A) < P(B)$

Th 1.2.10 (Bonferroni's inequality)

$$P(A \cap B) \geq P(A) + P(B) - 1$$

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$$

Th 1.2.11 (Results for dealing with a collection of sets)
If P is a probability function, then

1. $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ for any partition $C_1, C_2, \dots$
2. $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$ for any sets $A_1, A_2, \dots$ (Boole's inequality)

Proof for 2:
We first establish a disjoint collection $A_1^*, A_2^*, \ldots$, with the property that $\cup_{i=1}^{\infty} A_i^* = \cup_{i=1}^{\infty} A_i$.

$$A_1^* = A_1, \quad A_i^* = A_i \backslash \left(\bigcup_{j=1}^{i-1} A_j \right), i = 2, 3, \ldots$$

Therefore, we have

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

where the last equality follows since the A_i^* are disjoint. To see this, we write

$$\begin{aligned} A_i^* \cap A_k^* &= \left\{ A_i \backslash \left(\bigcup_{j=1}^{i-1} A_j \right) \right\} \cap \left\{ A_k \backslash \left(\bigcup_{j=1}^{k-1} A_j \right) \right\} (\text{definition of } A_i^*) \\ &= \left\{ A_i \cap \left(\bigcup_{j=1}^{i-1} A_j \right)^c \right\} \cap \left\{ A_k \cap \left(\bigcup_{j=1}^{k-1} A_j \right)^c \right\} (\text{definition of } \backslash \,) \\ &= \left\{ A_i \cap \bigcap_{j=1}^{i-1} A_j^c \right\} \cap \left\{ A_k \cap \bigcap_{j=1}^{k-1} A_j^c \right\} (\text{DeMorgan's Laws}) \end{aligned}$$

Now if $i > k$, the first intersection above will be contained in the set A_k^c, which will have an empty intersection with A_k. If $k > i$, the argument is similar. Further, by construction $A_i^* \subset A_i$, so $P(A_i^*) \leq P(A_i)$ and we have

$$\sum_{i=1}^{\infty} P(A_i^*) \leq \sum_{i=1}^{\infty} P(A_i)$$

Th 1.2.14 (Fundamental theorem of counting)
If a job consists of k seperate tasks, the i-th of which can be done in n_i ways, then the entire job can be done in $n_1 \times n_2 \times \cdots \times n_k$ ways.

Def 1.2.17 (Combination)
For nonnegative integers n and r, where $n \geq r$, we define the symbol $\binom{n}{r} = \frac{n!}{r!(n-r)!}$

Remark 1.2.18 (Number of possible arrangements of size r from n objects)

1. Ordered, without replacement: $\frac{n}{(n-r)!}$
2. Ordered, with replacement: n^r
3. Unordered, without replacement: $\binom{n}{r}$
4. Unordered, with replacement: $\binom{n+r-1}{r}$

Remark 1.2.19 (Methods for poker game)
If we wish to calculate probabilities for events that depend on the order, such as the probability of an ace in the first two cards, then we must use the ordered outcomes. If we want to calculate the probability of an event that does not depend on the order, we can use either the ordered or unordered sample space.

Remark 1.2.20 (Sampling with replacement)
Consider sampling $r = 2$ items from $n = 3$ items, with replacement.

Unordered	{1,1}	{2,2}	{3,3}	{1,2}	{1,3}	{2,3}
Ordered	(1,1)	(2,2)	(3,3)	(1,2),(2,1)	(1,3),(3,1)	(2,3),(3,2)
Probability	1/9	1/9	1/9	2/9	2/9	2/9

The formula for the number of outcomes in the unordered sample space is useful for enumerating the outcomes, but ordered outcomes must be counted to correctly calculate probabilities.

1.3 Conditional Probability and Independence

Def 1.3.2 (Conditional probability)
If A and B are events in S, and $P(B) > 0$, then the conditional probability of A given B, written $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Th 1.3.5 (Bayes' rule)
Let $A_1, A_2, \ldots$ be a partition of the sample space, and let B be any set. Then, for each $i = 1, 2, \ldots$,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)}$$

Def 1.3.7 (Independence)
Two events, A and B, are statistically independent if $P(A \cap B) = P(A)P(B)$

Th 1.3.9 (Independence)
If A and B are independent events, then the following pairs are also independent:

1. A and B^c
2. A^c and B
3. A^c and B^c

Remark 1.3.8 (Independence among more than two events)

The requirement $P(A \cap B \cap C) = P(A)P(B)P(C)$ is NOT a strong enough condition to garantee pairwise independence.

If A, B, C are pairwise independent, there might be $P(A \cap B \cap C) \neq P(A)P(B)P(C)$.

Def 1.3.12 (Mutually independent)
A collection of events $A_1, \dots, A_n$ are mutually independent if for any subcollection $A_{i_1}, \dots, A_{i_k}$, we have

$$P(\bigcap_{j=1}^{k} A_{i_j}) = \prod_{j=1}^{k} P(A_{i_j})$$

1.4 Random Variables

Def 1.4.1 (Random variables)
A random variable is a function from a sample space S into real numbers.

Remark 1.4.2 (Random variables)
In defining a random variable, we have also define a new sample space (the range of the random variable). We must check formally that our probability function, which is defined on the original sample space, can be used for the random variable.

Suppose we have a sample space $S = \{s_1, \dots, s_n\}$ with a probability function P, and we define a random variable X with range $\mathcal{X} = \{x_1, \dots, x_m\}$. We can define a probability function P_X on $\mathcal{X}$ in the following way.

$$P_X(X = x_i) = P(\{s_j \in S : X(s_j) = x_i\})$$

Such is also the case if $\mathcal{X}$ is countable. If $\mathcal{X}$ is uncountable, we define in the following way. For any set $A \subset \mathcal{X}$,

$$P_X(X \in A) = P(\{s \in S : X(s) \in A\})$$

1.5 Distribution Functions

Def 1.5.1 (Cummulative distribution function)
The cumulative distribution function or cdf of a random variable X, denoted by $F_X(x)$, is defined by

$$F_X(x) = P_X(X \leq x), \quad \text{for all } x$$

Th 1.5.3 (Properties of cdf)
The function $F(x)$ is a cdf if and only if the following three conditions hold:

1. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
2. $F(x)$ is a nondecreasing function of x
3. $F(x)$ is a right-continuous; that is, for every number x_0, $\lim_{x \downarrow x_0} F(x) = F(x_0)$

Def 1.5.7 (Continuous variable)
A random variable X is continuous if $F_X(x)$ is a continuous function of x. A random variable X is descrete if $F_X(x)$ is a step function of x.

Def 1.5.8 (Identically distributed)
The random variables X and Y are identically distributed if, for every set $A \in \mathcal{B}^1$, $P(X \in A) = P(Y \in A)$

Notice that two random variables that are identically distributed are not necessarily equal.

Th 1.5.10 (Identically distributed)
The following two statement are equivalent:

1. The random variables X and Y are identically distributed.
2. $F_X(x) = F_Y(x)$ for every x.

1.6 Density and Mass Functions

Def 1.6.1 (Probability mass function)
The Probability mass function (pmf) of a discrete random variable X is given by

$$f_X(x) = P(X = x) \quad \text{for all } x.$$

Def 1.6.3 (Probability density function)
The Probability density function or pdf, $f_X(x)$, of a continuous random variable X is the function that satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

Th 1.6.5 (Properties of pdf or pmf)
A function $f_X(x)$ is a pdf (or pmf) of a random variable X if and only if

1. $f_X(x) \geq 0$ for all x.
2. $\sum_x f_X(x) = 1$ (pmf) or $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$ (pdf).

1.7 Exercices

1.12, 1.18, 1.19, 1.23, 1.25, 1.27, 1.28, 1.45

Some useful formula:
$\sum_{x=0}^{n} \binom{n}{x}^2 = \binom{2n}{n}$,
$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0$, which can be proved using $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$,
$\sum_{k=1}^{n} k\binom{n}{k} = n2^{n-1}$,
$\sum_{k=1}^{n} (-1)^{k+1} k\binom{n}{k} = 0$
(Stirling's formula) $n! \approx \sqrt{2\pi} n^{n+1/2} e^{-n}$

Chapter 2 Transformation and Expectation

2.1 Distributions of Functions of a Random Variable

Eg 2.1.1 (Binomial transformation)
A discrete random variable X has a binomial distribution if its pmf if of the form

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

where n is a positive integer and $0 \leq p \leq 1$.

Consider the random variable $Y = g(X)$, where $g(x) = n - x$, that is, $Y = n - X$. Here $\mathcal{X} = \{0, 1, \ldots, n\}$ and $\mathcal{Y} = \{y : y = g(x), x \in \mathcal{X}\} = \{0, 1, \ldots, n\}$. For any $y \in \mathcal{Y}$, $n - x = g(x) = y$ if and only if $x = n - y$. Thus, $g^{-1}(y)$ is the single point $x = n - y$, and

$$\begin{aligned} f_Y(y) &= \sum_{x \in g^{-1}(y)} f_X(x) \\ &= f_X(n-y) \\ &= \binom{n}{n-y} p^{n-y} (1-p)^{n-(n-y)} \\ &= \binom{n}{y} (1-p)^y p^{n-y} \end{aligned}$$

Eg 2.1.2 (Uniform transformation)
Suppose X has a uniform distribution, on the interval $(0, 2\pi)$, that is

$$f_X(x) = \begin{cases} 1/2\pi & 0 < x < 2\pi \\ 0 & \text{otherwise} \end{cases}$$

Consider $Y = \sin^2(X)$. Then

$$P(Y \leq y) = P(X \leq x_1) + P(x_2 \leq X \leq x_3) + P(X \geq x_4)$$

From the symmetry of the function $\sin^2(x)$, and the fact that X has a uniform distribution, we have

$$P(Y \leq y) = 2P(X \leq x_1) + 2P(x_2 \leq X \leq \pi)$$

where x_1 and x_2 are two solutions to $\sin^2(x) = y$, $(0 \leq x \leq \pi)$.

Th 2.1.3 (cdf of transformation)
Let X have cdf $F_X(x)$, let $Y = g(X)$, and let $\mathcal{X}$ and $\mathcal{Y}$ be defined as $\mathcal{X} = \{x : f_X(x) > 0\}$, $\mathcal{Y} = \{y : y = g(x) \text{ for some } x \in \mathcal{X}\}$.

1. If g is an increasing function on $\mathcal{X}$, $F_Y(y) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.
2. If g is an decreasing function on $\mathcal{X}$, and X is a continuous random variable, $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.

Th 2.1.5 (pdf of transformation)
Let X have pdf $f_X(x)$ and let $Y = g(X)$, where g is a monotone function. Let $\mathcal{X}$ and $\mathcal{Y}$ be defined as **Th 2.1.3**. Supoose that $f_X(x)$ is continuous on $\mathcal{X}$ and that $g^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$. Then the pdf of Y is given by

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|, & y \in \mathcal{Y} \\ 0, & \text{otherwise} \end{cases}$$

Th 2.1.8 (General case of transformation)
Let X have pdf $f_X(x)$, let $Y = g(X)$, and define the sample space $\mathcal{X}$ as **Th 2.1.3**. Suppose there exists a partition, $A_0, A_1, \ldots, A_k$, of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further, suppose there exists functions $g_1(x), \ldots, g_k(x)$, defined on $A_1, \ldots, A_k$, respectively, satisfying

1. $g(x) = g_i(x)$, for $x \in A_i$,
2. $g_i(x)$ is monotone on A_i,
3. the set $\mathcal{Y} = \{y : y = g_i(x) \text{ for some } x \in A_i\}$ is the same for each $i = 1, \ldots, k$,
4. $g_i^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$, for each $i = 1, \ldots, k$.

Then

$$f_Y(y) = \begin{cases} \sum_{i=1}^k f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right|, & y \in \mathcal{Y} \\ 0, & \text{otherwise} \end{cases}$$

Th 2.1.10 (Probability integral transformation)
Let X have continuous cdf $F_X(x)$ and define the random variable Y as $Y = F_X(X)$. Then Y is uniformly distributed on $(0, 1)$, that is, $P(Y \le y) = y, 0 < y < 1$.

Proof: For $Y = F_X(X)$, we have, for $0 < y < 1$,

$$\begin{aligned} P(Y \le y) &= P(F_X(X) \le y) \\ &= P(F_X^{-1}[F_X(X)] \le F_X^{-1}(y)) && (F_X^{-1} \text{ is increasing}) \\ &= P(X \le F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) && (\text{definition of } F_X) \\ &= y && (\text{continuity of } F_X) \end{aligned}$$

At the endpoints we have $P(Y \le y) = 1$ for $y \ge 1$ and $P(Y \le y) = 0$ for $y \le 0$, showing that Y has a uniform distribution.

The reasoning behind $P(F_X^{-1}[F_X(X)] \le F_X^{-1}(y)) = P(X \le F_X^{-1}(y))$ is subtle and deserves additional attention. If F_X is not strictly increasing, it may be that $F_X^{-1}(F_X(x)) \ne x$. Suppose $F_X^{-1}(y) = \inf\{x : F_X(x) \ge y\}$. If F_X is flat when $x \in [x_1, x_2]$. $F_X^{-1}(F_X(x)) = x_1$ for any x in this interval. Even in this case, the probability equation holds, and the flat cdf denotes a region of 0 probability $P(x_1 < X \leq x) = F_X(x) - F_X(x_1) = 0$.

2.2 Expected Values

Def 2.2.1 (Expected value)
The expected value or mean of a random variable $g(X)$, denoted by $E\,g(X)$, is

$$E\,g(X) = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x)\,dx, & \text{if } X \text{ is continuous} \\ \sum_{x\in\mathcal{X}} g(x) f_X(x) = \sum_{x\in\mathcal{X}} g(x) P(X=x), & \text{if } X \text{ is discrete,} \end{cases}$$

provided that the integral or sum exists. If $E|g(X)| = \infty$, we say that $E\,g(X)$ does not exist.

Th 2.2.5 (Properties of expectation)
Let X be a random variable and let a, b and c be constants. Then for any functions $g_1(x)$ and $g_2(x)$ whose expectations exist,

1. $E(ag_1(x) + bg_2(x) + c) = aE(g_1(x)) + bE(g_2(x)) + c$.
2. If $g_1(x) \geq 0$ for all x, then $E\,g_1(X) \geq 0$.
3. If $g_1(x) \geq g_2(x)$ for all x, then $E\,g_1(X) \geq E\,g_2(X)$.
4. If $a \leq g_1(x) \leq b$ for all x, then $a \leq E\,g_1(X) \leq b$.

Chapter 4 Multiple Random Variables

4.1 Joint and Marginal Distributions

Def 4.1.1 (Random vector)
An n-dimensional random vector is a funciton from a sample space S into $\mathcal{R}^n$, n-dimensional Educlidean space.

Def 4.1.3 (Joint pmf)
Let (X, Y) be a discrete bivariate random vector. Then the function $f(x, y)$ from $\mathcal{R}^2$ into $\mathcal{R}$ defined by $f(x, y) = P(X = x, Y = y)$ is called the joint probability mass function or joint pmf of (X, Y).

Def 4.1.6 (Marginal pmf)
Let (X, Y) be a descrete bivariate random vector with joint pmf $f_{X,Y}(x, y)$. Then the marginal pmfs of X and Y, $f_X(x) = P(X = x)$ and $f_Y(y) = P(Y = y)$, are given by

$$F_X(x) = \sum_{y\in\mathcal{R}} f_{X,Y}(x,y) \quad \text{and} \quad F_Y(y) = \sum_{x\in\mathcal{R}} f_{X,Y}(x,y)$$

Def 4.1.10 (Joint pdf)
A function $f(x, y)$ from $\mathcal{R}^2$ into $\mathcal{R}$ is called a joint probability density function or joint pdf of the

continuous bivarite random vector (X, Y) if, for every $A \subset \mathcal{R}^2$,

$$P((X, Y) \in A) = \int_A \int f(x, y) \, dx \, dy.$$

4.2 Conditional Distributions and Independence

Def 4.2.1 (Conditional pmf)
Let (X, Y) be a descrete bivariate random vector with joint pmf $f(x, y)$ and marginal pmfs $f_X(x)$ and $f_Y(y)$. For any x such that $P(X = x) = f_X(x) \geq 0$, the conditional pmf of Y given that $X = x$ is the function of y denoted by $f(y|x)$ and defined by

$$f(y|x) = P(Y = y|X = x) = \frac{f(x, y)}{f_X(x)}.$$

Def 4.2.3 (Conditional pdf)
Let (X, Y) be a continuous bivariate random vector with joint pdf $f(x, y)$ and marginal pdfs $f_X(x)$ and $f_Y(y)$. For any x such that $f_X(x) > 0$, the conditional pdf of Y given that $X = x$ is the function of y denoted by $f(y|x)$ and defined by

$$f(y|x) = \frac{f(x, y)}{f_X(x)}.$$

Def 4.2.5 (Independent random variables)
Let (X, Y) be a bivariate random vector with joint pdf or pmf $f(x, y)$ and marginal pdfs or pmfs $f_X(x)$ and $f_Y(y)$. Then X and Y are called independent random variables if, for every $x \in \mathcal{R}$ and $y \in \mathcal{R}$,

$$f(x, y) = f_X(x) f_Y(y)$$

Lemma 4.2.7 (Independent random variables)
Let (X, Y) be a bivariate random vector with joint pdf or pmf $f(x, y)$. Then X and Y are independent random variables if and only if there exist functions $g(x)$ and $h(y)$ such that, for every $x \in \mathcal{R}$ and $y \in \mathcal{R}$,

$$f(x, y) = g(x) h(y)$$

Th 4.2.10 (Properties of independent variables)
Let X and Y be independent random variables.

1. For any $A \subset \mathcal{R}$ and $B \subset \mathcal{R}$, $P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$; that is, the events$\{X \in A\}$ and $\{Y \in B\}$ are independent events.
2. Let $g(x)$ be a function only of x and $h(y)$ only be a function of y. Then $E(g(X)h(Y)) = E(g(X)) \cdot E(h(Y))$

Th 4.2.12 (Independent r.v.s' moment generating functions)
Let X and Y be independent random variables with moment generating functions $M_X(t)$ and $M_Y(t)$. Then the moment generating function of the random variable $Z = X + Y$ is given by $M_Z(t) = M_X(t)M_Y(t)$.

Th 4.2.14 (Independent normal random variables)
Let $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\gamma, \tau^2)$ be independent normal random variables. Then the random variable $Z = X + Y \sim N(\mu + \gamma, \sigma^2 + \tau^2)$.

4.3 Bivariate Transformation

Th 4.3.2 (Independnet Poisson)
If $X \sim \mathrm{Poisson}(\theta)$ and $Y \sim \mathrm{Poisson}(\lambda)$ and X and Y are independent, then $X + Y \sim \mathrm{Poisson}(\theta + \lambda)$.

Th 4.3.5 (Independent)
Let X and Y be independent random variables. Let $g(x)$ be a function only of x and $h(y)$ be a function only of y. Then the random variables $U = g(X)$ and $V = h(Y)$ are independent.

proof:
Assuming U and V are continuous r.v.. For any $u \in \mathcal{R}$ and $v \in \mathcal{R}$, define

$$A_u = \{x : g(x) \le u\} \quad \text{and} \quad B_v = \{y : h(y) \le v\}.$$

Then the joint cdf of (U, V) is

$$\begin{aligned} F_{U,V}(u, v) &= P(U \le u, V \le v) && \text{(definition of cdf)} \\ &= P(X \in A_u, Y \in B_v) && \text{(definition of } U \text{ and } V\text{)} \\ &= P(X \in A_u)P(Y \in B_v) && \text{(Th 4.2.10)} \end{aligned}$$

The joint pdf of (U, V) is

$$\begin{aligned} f_{U,V}(u, v) &= \frac{\partial^2}{\partial u \partial v} F_{U,V}(u, v) \\ &= \left(\frac{d}{du} P(X \in A_u)\right)\left(\frac{d}{dv} P(Y \in B_v)\right), \end{aligned}$$

By Th 4.2.7, U and V are independent.