STA257 Probability and Statistics I

Gun Ho Jang

Update on May 14, 2018

Note: This note is prepared for STA257. There might be numerous fault arguments/statements/typos. If you spot one, please contact the instructor or you may look up references which may contain errors too.

Random Variables and Distributions

Random Variables and Discrete Distributions

Definition 10. A real-valued function X on the sample space S is called a *random variable* if the probability of X is well-defined, that is, $\{s \in S : X(s) \leq r\}$ is an event for each $r \in \mathbb{R}$.

Most random functions in this course are random variables. Hence any functions on sample space will be treated as random variables hereafter.

Example 30. Suppose a balanced coin is tossed 5 times. The number of head in each toss is a random variable.

The sample space is $S = \{H, T\}^5$ and $X : S \to \mathbb{R}$ is defined as the number of heads like $X(TTHTH) = 2$.

Note. For any Borel set B in $\mathbb{R}$, an event $X \in B$ is defined as $\{s \in S : X(s) \in B\}$ and often denoted by $\{X \in B\}$ or $(X \in B)$. The corresponding probability is

$$P(X \in B) = P(\{s \in S : X(s) \in B\}).$$

Events are often in the forms of $\{X \leq a\}, \{X < a\}, \{X \geq a\} = \{X < a\}^c, \{X > a\} = \{X \leq a\}^c, \{a < X \leq b\} = \{X > a\} \cap \{X \leq b\} = \{X \leq a\}^c \cap \{X \leq b\}$ for any real numbers $a < b \in \mathbb{R}$.

Definition 11 (Borel sets in $\mathbb{R}$)**.** The collection of all Borel sets $\mathcal{B}$ in $\mathbb{R}$ is the smallest collection satisfying the followings
(a) [closure under complement] For any $B \in \mathcal{B}$, also $B^c \in \mathcal{B}$.
(b) [closure under countable union] For any $B_1, B_2, \ldots \in \mathcal{B}$, $\bigcup_{j=1}^{\infty} B_j \in \mathcal{B}$.
(c) $(a, b] \in \mathcal{B}$ for any $a < b \in \mathbb{R}$.

Roughly speaking, Borel sets are induced events from intervals in $\mathbb{R}$. The collection $\mathcal{B}$ is often called the Borel σ-field of $\mathbb{R}$ which contains uncountably many subsets.

Lemma 15. If $|X(S)| < \infty$ and $(X = r)$ is event for any $r \in X(S)$, then X is a random variable.

Note. Sometimes the space of objects becomes a sample space and a feature from each object becomes random variable. For example, S can be a collection of all U of T students and $X(s)$ is the student s's physical height for each $s \in S$.

Example 31. When a fair die is thrown twice, let X, Y, Z be the sum, minimum and maximum. Then all three are random variables are derived from one sample space. The sample space is $S = \{1, \ldots, 6\}^2 = \{(i,j) : 1 \le i, j \le 6\}$ and $X(i,j) = i+j, Y(i,j) = \min(i,j), Z(i,j) = \max(i,j)$ are defined on S.

Example 32 (Function of random vairables)**.** If X and Y are random variables, then so are aX for $a \in \mathbb{R}$, $X+Y$, and XY.
Suppose $a > 0$, for any $r \in \mathbb{R}$, $P(aX > r) = P(X > r/a)$ is defined. Hence $(aX > r)$ is an event, which concludes aX is a random variable.
Note that $(X+Y > r) = \bigcup_{q \in \mathbb{Q}} ((X > q) \cap (Y > r-q))$ which is a countable union of events. Thus $X+Y$ is a random variable.

Note. The next statement can be proved using mathematical theory. "If $h : \mathbb{R}^2 \to \mathbb{R}$ is continuous, then $h(X,Y)$ is also a random variable whenever X and Y are random variables."

Using the theorem we can prove $X+Y$ and XY are random variables. To prove $X+Y$ is a random variable we need the fact that if a function $h : \mathbb{R}^2 \to \mathbb{R}$ is continuous, then $h(X,Y)$ is a random variable. By noting that $h_+(x,y) = x+y$ and $h_\times(x,y) = xy$ are continuous, $X+Y = h_+(X,Y)$ and $XY = h_\times(X,Y)$ are random variables.

Exercise 11. Let $Y, X, X_1, X_2, \ldots$ be random variables. Show that $X_1 + \cdots + X_n$, $\sup_n X_n$, $\inf_n X_n$, $\limsup_n X_n$, and $\liminf_n X_n$ are random variables. [Hint: $(\sup_n X_n > r) = \cup_n (X_n > r)$]

Definition 12. The *distribution* of X is the collection of all probabilities of all events induced by X, that is, $(B, P(X \in B))$. Two random variables X and Y are said to be *identically distributed* if they have the same distribution.

To show X and Y having the same distribution, we need to check for any event B on $\mathbb{R}$, $P(X \in B) = P(Y \in B)$. Since all Borel sets on $\mathbb{R}$ are induced by intervals, it is enough to prove $P(a < X \le b) = P(a < Y \le b)$ for any $a < b \in \mathbb{R}$. Even $P(X \le a) = P(Y \le a)$ for any $a \in \mathbb{R}$ guarantees that X and Y are identically distributed.

Example 33. Let X, Y be the numbers of heads and tails in 10 fair coin tosses. The sample space is $S = \{H, T\}^{10}$. For each outcome $s \in S$, the probability is the same. Hence $P(\{s\}) = 1/2^{10}$. The possible values of X, Y are $0, 1, \ldots, 10$. Then $P(X = x)$ is the number outcomes having x heads multiplied by the probability $1/2^{10}$, that is, $P(X = x) = \binom{10}{x} \times 1/2^{10}$. Similarly $P(Y = y) = \binom{10}{y} \times 2^{-10}$. Hence $P(X = x) = \binom{10}{x}/2^{10} = P(Y = x)$ and X and Y have the same distribution.

In the above example, the size of sample space is $|S| = 2^1 0 = 1024$ and the number of events is $2^{|S|} = 2^{1024} = 1.79763 \times 10^{308}$ which is extremely big number. While the number of possible values of X was 11 $(0, 1, \ldots, 10)$ and the events of X is $2^{|X(S)|} = 2^{11} = 2048$ which is manageable compare to the number of events from S. By using random variables, computation of probability gets easier in general.

Definition 13. A random variable X is said to be *discrete* if the number of its values is finite or countably many.

In some text books, a random variable X is said to be *discrete* if $P(X = r) = 0$ or $P(X = r) > 0$ for any real number $r \in \mathbb{R}$ and $\sum_{r \in \mathbb{R}} P(X = r) = 1$. Such random variable takes at most countably many values $x_1, x_2, \ldots$ such that $P(X = x_i) > 0$ and $\sum_{i=1}^{\infty} P(X = x_i) = 1$.

There are *continuous* random variables as a counter concept.

Definition 14. The *probability mass function* (p.m.f.) of a discrete random variable X is

$$\mathrm{pmf}_X(x) = P(X = x)$$

for any possible value of $x \in X(S)$.

Some authors prefer "probability function" rather than "probability mass function."

Example 34. Let X be the number of head in n tosses of a balanced coin. Then X takes values $0, 1, \ldots, n$ and

$$\mathrm{pmf}_X(x) = \begin{cases} \binom{n}{x}/2^n & \text{if } x = 0, 1, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 16. Let X be a discrete random variable. Then the set of x having $P(X = x)$ is at most countable.

Proof. It becomes trivial by accepting Definition 13. When *discrete* is defined as $\sum_{r \in \mathbb{R}} P(X = r) = 1$. Then the followings is a proof of the theorem.

Define $A_n = \{x : P(X = x) \geq 1/n\}$. Then $|A_n| \leq n$ and $A = \{x : P(X = x) > 0\} = \cup_n A_n$ is countable union of finite sets. Hence A has at most countably many elements. $\square$

Theorem 17. Let f be the probability mass function of a discrete random variable X. The set of possible values of X is $X(S) = \{x_1, x_2, \ldots\}$. For $x \notin X(S)$, $f(x) = 0$ and for $x_i \in X(S)$, $f(x_i) \geq 0$ and $\sum_{i=1}^{\infty} f(x_i) = 1$.

Theorem 18. Let $X(S) = \{x_1, x_2, \ldots\}$ be the set of possible values of a discrete random variable X. Then for any subset A of $\mathbb{R}$,

$$P(X \in A) = \sum_{x_i \in A} P(\{x_i\}) = \sum_{x_i \in A} \mathrm{pmf}_X(x_i).$$

Definition 15. A random variable X taking values 0 and 1 with $P(X = 1) = p$ and $P(X = 0) = 1 - p$ for some $p \in [0, 1]$ is called a *Bernoulli* random variable with success probability p and often denoted by $X \sim \mathrm{Bernoulli}(p)$.

Definition 16. Let $\mathcal{X}$ be a non-empty finite set. A random variable X taking values in $\mathcal{X}$ with equal probability is called a *uniform* random variable on $\mathcal{X}$ and denoted by $X \sim \mathrm{uniform}(\mathcal{X})$.

The probability mass function of $X \sim \mathrm{uniform}(\mathcal{X})$ is

$$\mathrm{pmf}_X(x) = \begin{cases} \frac{1}{|\mathcal{X}|} & \text{if } x \in \mathcal{X}, \\ 0 & \text{otherwise.} \end{cases}$$

Exercise 12. Show that $X \sim \mathrm{Bernoulli}(1/2)$ and $Y \sim \mathrm{uniform}(\{0, 1\})$ have the same probability mass function.

Definition 17. A random variable X is called a *binomial* random variable if it has the same distribution as Z which is the number of success in n independent trails with with success probability p, and denoted by $X \sim \text{binomial}(n,p)$.

The probability mass function of $X \sim \text{binomial}(n,p)$ is

$$\text{pmf}_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } n = 0, \ldots, n, \\ 0 & \text{otherwise.} \end{cases}$$

Continuous Distributions

Definition 18. A random variable X is said to be *continuous* if the probability of each interval $[a,b]$ is of the form

$$P(a < X \leq b) = \int_a^b f(x)\, dx$$

where $a < b \in \mathbb{R}$ and f is a non-negative function on $\mathbb{R}$. Such function f is called a *probability density function* (p.d.f.) of X.

A probability density function is often called as a *density function* by some authors.

Definition 19. A random variable X defined on (a,b) for finite real numbers $a < b$ satisfying $P(c < X \leq d) = (d-c)/(b-a)$ for any c,d such that $a \leq c \leq d \leq b$ is called a *uniform* random variable on (a,b) which is denoted by $X \sim \text{uniform}(a,b)$.

Theorem 19. The probability density function of $X \sim \text{uniform}(a,b)$ is

$$\text{pdf}_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The definition of uniform distribution implies

$$P(c \leq X \leq d) = \frac{d-c}{b-a} = \int_c^d \frac{1}{b-a}\, dx = \frac{1(a \leq x \leq b)}{b-a}\, dx$$

where $1(\cdot)$ is an indicator function having value 1 if the condition in $(\cdot)$ is satisfied and value 0 otherwise. Hence the density of X is given by

$$\text{pdf}_X(x) = \frac{1(a \leq x \leq b)}{b-a}.$$

Recall the second fundamental theorem of calculus: For a continuous integrable function f defined on (a,b), define

$$F(x) = \int_a^x f(z)\, dz.$$

Then $F'(x) = \frac{d}{dx} F(x) = f(x)$.

Thus for any $a < x < b$,

$$\frac{d}{dx}P(a < X \leq x) = \frac{d}{dx}\int_a^x \mathrm{pdf}_X(z)\, dz = \mathrm{pdf}_X(x)$$

Hence the density at x is given by

$$\mathrm{pdf}_X(x) = \frac{d}{dx}P(a < X \leq x) = \frac{d}{dx}\frac{x-a}{b-a} = \frac{1}{b-a}.$$

Also when $x < a$ or $x > b$, the density becomes $\mathrm{pdf}_X(x) = 0$. $\square$

Example 35. Consider a random variable X having density

$$\mathrm{pdf}_X(x) = \begin{cases} cx & \text{for } 0 < x < 4, \\ 0 & \text{otherwise.} \end{cases}$$

Find value of c and compute $P(1 < X \leq 2)$ and $P(X > 2)$.

Density must be non-negative and integrated to 1 (Axiom 2 of the probability). Hence $c \geq 0$ and $1 = \int_0^4 \mathrm{pdf}_X(x)\, dx = \int_0^4 cx\, dx = cx^2/2|_0^4 = 8c$ implies $c = 1/8$.

Then $P(1 < X \leq 2) = \int_1^2 x/8\, dx = x^2/16|_1^2 = 3/16$ and $P(X > 2) = \int_2^4 x^2/8\, dx = x^2/16|_2^4 = 3/4$.

Example 36 (A random variable which is neither discrete nor continuous)**.** Let X be the size of molecules which is supposed to follow Uniform$(0, 100)$. Let Y be the size measured of X through a microscope which can detect molecules having size at least 10 units. It is easy to see that $Y = 0$ if $X < 10$ and $= X$ if $X \geq 10$. Then $P(Y = 0) = P(X < 10) = \int_0^{10} 1/100\ dz = 10/100 = 0.1$ and for $10 < x < 100$, $P(Y \leq x) = P(X \leq x) = \int_0^x 1/100\, dz = x/100$.

Examples of Random Variables

Definition 20. Consider an independent Bernoulli trial with success probability p. The number of trials until the first success is called a *geometric* distribution with parameter p, denoted by geometric(p), and the number of trials until k-th success is called a *negative binomial* distribution with parameter k and p, denoted by neg-bin(k, p).

The geometric random variable $X \sim$ geometric(p) has probability mass function as $\mathrm{pmf}_X(n) = (1-p)^{n-1}p$ for natural numbers n because there will be $n-1$ failure until the first success. Similarly the probability mass function of $Y \sim$ neg-bin(k, p) is $\mathrm{pmf}_Y(n) = \binom{n-1}{k-1}(1-p)^{n-k}p^k$ considering there are $k-1$ success among the first and $(n-1)$-st trial and the last trial must be success.

Consider a jar containing n balls of which r are black and the remainder $n - r$ are white. The random variable X is the number of black balls when m balls are drawn without replacement. The probability of k black balls are drawn is

$$\mathrm{pmf}_X(k) = \begin{cases} \binom{r}{k}\binom{n-r}{m-k}\Big/\binom{n}{m} & \text{if } k = 0, \ldots, \min(r, m) \\ 0 & \text{otherwise.} \end{cases}$$

Such distribution is called a *hypergeometric* distribution.

The Poisson distribution is related to the rate of certain events occurrence. For example, the number of cars passing per unit time on a popular street has a Poisson(μ) distribution if cars are passing at the rate of μ cars per unit time, no two cars are passing at the same time, the cars passing rates are proportional to the length of time intervals.

When the unit time is divided into k equal length time, that is, $1/k$ unit time. On each interval, cars are expected to pass at the rate of μ/k. Say $X_{k,1}, \ldots, X_{k,k}$ are the number of cars passing in sub intervals respectively. When k is big enough, no two cars are expected to pass within a sub interval, hence assume $P(X_{k,j} \geq 2) = o(1/k)$ as $k \to \infty$. Also assume passing one car is $P(X_{k,j} = 1) = \mu/k + o(1/k)$. Then $P(X_{k,j} = 0) = 1 - P(X_{k,j} = 1) - P(X_{k,j} \geq 2) = 1 - \mu/k + o(1/k)$ and

$$P(X = 0) = P(X_{k,1} = \cdots = X_{k,k} = 0) = [1 - \mu/k + o(1/k)]^k = \exp(k \times (-\mu/k + o(1/k))) = \exp(-\mu + o(1)) \to \exp(-\mu)$$

as $k \to \infty$. Let E_k be the event that there exists a subinterval in which at least two cars passing. Then

$$P(E_k) = P(\overset{k}{\underset{j=1}{\cup}} (X_{k,j} \geq 2) \leq \sum_{j=1}^{k} P(X_{k,j} \geq 2) = kP(X_{k,1} \geq 2) = o(1).$$

Let Y_k be the number of subintervals in which only one car passing. It is easy to see that $Y_k \sim \text{binomial}(k, P(X_{k,1} = 1))$. Then $P(X = n) = P(X = n, E_k^c) + P(X = n, E_k) = P(Y_k = n) + P(X = n, E_k)$ and it implies $|P(X = n) - P(Y_k = n)| \leq P(E_k) = o(1)$, that is, $P(X = n) = P(Y_k = n) + o(1) = \binom{k}{n}(\mu/k + o(1/k))^n(1 - \mu/k + o(1/k))^{k-n} + o(1) = \frac{k!}{n!(k-n)!}k^{-n}(\mu)^{n+o(1)}\exp(-\mu/k \times (k-n) + o(1)) = \mu^n/n! \exp(-\mu) + o(1) \to e^{-\mu}/n! \times \mu^n$. Thus $P(X = n) = e^{-\mu}\mu^n/n!$ for $n = 0, 1, 2, \ldots$.

Let Y be the time to the first car passing which is also interesting. For example, consider the probability of the first car passed between time 0 and y. Let Z be the number of cars passed between time 0 and y units. Then $Z \sim \text{Poisson}(y\mu)$ and $P(Y \leq y) = P(\text{at least one car passed at time } y) = P(Z \geq 1) = 1 - P(Z = 0) = 1 - e^{-y\mu}$. Hence the probability density of Y is $\text{pdf}_Y(y) = e^{-\mu y}$ for $y \geq 0$ and 0 otherwise because

$$P(Y \leq y) = 1 - e^{-\mu y} = \int_0^y \mu e^{-\mu w}\, dw.$$

The probability density of Y is decreasing exponentially. As a note, $W \sim \text{exponential}(\lambda)$ has a density $\text{pdf}_W(w) = \lambda e^{-\lambda w} 1(w > 0)$. Hence $Y \sim \text{exponential}(\mu)$.

Densities decreasing slower than exponential distributions are also interested in many statistical models. One of the most common densities are proportional to $x^{\alpha-1}e^{-\beta x}1(x > 0)$ which is a multiplication of exponential density and a high order power function. Using change of variable $y = \beta x$ or $dy = \beta dx$, the integral becomes

$$\int_0^\infty x^{\alpha-1}e^{-\beta x}\, dx = \int_0^\infty (y/\beta)^{\alpha-1}e^{-y}\, dy/\beta = \beta^{-\alpha}\int_0^\infty y^{\alpha-1}e^{-y}\, dy = \beta^{-\alpha}\Gamma(\alpha).$$

Hence a family of distributions having the probability density functions of the form

$$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x}1(x > 0)$$

is called *gamma* distribution with parameters α, β which is denoted by $X \sim \text{gamma}(\alpha, \beta)$.

It is easy to see that exponential(λ) and gamma($1, \lambda$) have the same densities.

One of the most common continuous distributions in statistics is a normal distribution or a gaussian distribution given by

$$\mathrm{pdf}_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)$$

which is denoted by $Z \sim N(\mu, \sigma^2)$ for $\mu \in \mathbb{R}, \sigma > 0$.

The Cumulative Distribution Function

The distribution contains all probability information, that is, the probability of all possible events. When $|X(S)| = n$, the distribution must be described to 2^n events. It is a ridiculously big number. While probability mass function requires probabilities of n data points which can generate the distribution efficiently.

Similarly the distribution of a continuous random variable can be determined by the probability density function.

What can we do for random variables containing both discrete and continuous parts?

Probability mass/density functions may not exists. Hence a new aspect is required.

Definition 21. The *(cumulative) distribution function* of a random variable X is the function

$$\mathrm{cdf}_X(x) = F_X(x) = P(X \le x)$$

for $-\infty < x < \infty$.

Example 37. The distribution function of $X \sim \mathrm{Bernoulli}(p)$ is given by

$$\mathrm{cdf}_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1-p & \text{if } 0 \le x < 1, \\ 1 & \text{if } x \ge 1. \end{cases}$$

Theorem 20 (Properties of distribution functions)**.** Let F be a distribution function. Then (a) F is nondecreasing,
(b) $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$,
(c) F is right continuous, that is, $\lim_{y\searrow x} F(y) = F(x)$,
(d) $F(x-) \coloneqq \lim_{y\nearrow x} F(y) = P(X < x)$,
(e) $P(X = x) = F(x) - F(x-)$.

Proof. (a) Since $\{X \le x\} \subset \{X \le y\}$ for all $x \le y$, $F(x) = P(X \le x) \le P(X \le y) = F(y)$.
(b) $\lim_{x\to\infty} F(x) = \lim_{x\to\infty} P(X \le x) = P(X < \infty) = P(\Omega) = 1$ and $\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} P(X \le x) = P(X \le -\infty) = P(\emptyset) = 0$.
(c) For $y > x$, $F(y) = P(X \le y) = P(X \le x) + P(x < X \le y) = F(x) + P(X \in (x, y]) \to F(x)$ as $y \searrow x$ (Note that $\lim_{y\searrow x}(x, y] = (x, x] = \emptyset$).
(d) $F(x-) = \lim_{y\nearrow x} F(y) = \lim_{y\nearrow x} P(X \in (-\infty, y]) = P(X \in (-\infty, x)) = P(X < x)$.
(e) $P(X = x) = P(X \le x) - P(X < x) = F(x) - F(x-)$. $\square$

Theorem 21. Show that if a real function F satisfies (a)-(c) in the above properties, then it is a distribution function of a random variable.

Sketch proof. Assume $S = [0, 1]$ and $P([0, a]) = a$ for $a \in S$. Define $W : S \to \mathbb{R}$ so that $W(s) = \inf\{t : F(t) \ge s\}$. Obviously $\{W \le t\} = \{s \in S : W(s) \le t\} = \{s \in S : s \le F(t)\}$. Then, the distribution function of W is $\mathrm{cdf}_W(t) = P(W \le t) = P(\{s : s \le F(t)\}) = P([0, F(t)]) = F(t)$. Hence F is the distribution function of W. $\square$

Example 38 (Constant random variables)**.** For a real number $c \in \mathbb{R}$, let $X_c \equiv c$, that is, $X_c(s) = c$ for all $s \in S$. Noting that $P(X_c \le x) = 0$ if $x < c$, and $= 1$ if $x \ge c$, X_c is a random variable having distribution function given by

$$F_{X_c}(x) = \begin{cases} 0 & \text{if } x < c, \\ 1 & \text{if } x \ge c. \end{cases}$$

Example 39 (Indicator functions)**.** For any event A define $1_A : S \to \mathbb{R}$ such that

$$1_A(s) = \begin{cases} 1 & \text{if } s \in A, \\ 0 & \text{if } s \notin A. \end{cases}$$

Then $P(1_A \le x) = 0$ if $x < 0$, $= P(A)$ if $0 \le x < 1$, and $= 1$ if $x \ge 1$. Hence 1_A is a random variable.

Definition 22. The p-quantile of a random variable X is x such that $P(X \le x) \ge p$ and $P(X \ge x) \ge 1 - p$.

Exercise 13. Suppose the distribution function F is continuous. Show that p-quantile is $F^{-1}(p)$.

Example 40. The distribution function of $X \sim \text{uniform}[a, b]$ is

$$\mathrm{cdf}_X(x) = \begin{cases} 0 & \text{if } x < a, \\ (x - a)/(b - a) & \text{if } a \le x < b, \\ 1 & \text{if } x \ge b. \end{cases}$$

Definition 23. The *median*, *lower quartile*, *upper quartile* are 0.5-, 0.25-, 0.75-quantile. The *inter quartile range* (IQR) is the difference between upper and lower quartile.

Bivariate Distributions

Definition 24. The joint (or bivariate) *distribution* of two random variables X and Y is the collection of all possible probabilities, that is, $P((X, Y) \in B)$ where B is a Borel set in $\mathbb{R}^2$.

If (X, Y) is discrete as a pair, the it has a joint probability mass function. That is,

$$\begin{aligned} \mathrm{pmf}_{X,Y}(x, y) &= P((X, Y) = (x, y)) = P(X = x, Y = y) \\ &= P(\{s : X(s) = x, Y(s) = y)\}) \end{aligned}$$

Joint probability mass function $\mathrm{pmf}_{X,Y}(x, y)$ satisfies $\mathrm{pmf}_{X,Y}(x, y) \ge 0$ for any $x, y \in \mathbb{R}$ and there are $(x_1, y_1), (x_2, y_2), \ldots$ such that $\mathrm{pmf}_{X,Y}(x_n, y_n) > 0$ with $\sum_{n=1}^{\infty} \mathrm{pmf}_{X,Y}(x_n, y_n) = 1$.

Example 41. Consider U of T students. Let X and Y be the gender of student and the primary program the student is in. Then (X, Y) is a bivariate discrete random vector.

Definition 25. Two random variables X and Y are jointly continuously distributed if and only if there exists a non-negative function f such that for any Borel set B in $\mathbb{R}^2$

$$P((X,Y) \in B) = \iint\limits_B f(x,y)\, dx\, dy.$$

Such function f is called a joint density function of (X, Y).

Joint density functions satisfies $\mathrm{pdf}_{X,Y}(x,y) \geq 0$ and $\int\int \mathrm{pdf}_{X,Y}(x,y)\, dx\, dy = 1$.

Example 42. Consider random variables X and Y having density of the form

$$\mathrm{pdf}_{X,Y}(x,y) = 1(x^2 \leq y \leq 1)cx^2y.$$

Find the appropriate value $c > 0$ and compute $P(X \geq Y)$.

Axiom 2 implies

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1(x^2 \leq y \leq 1)cx^2y\, dy\, dx = \int_{-1}^{1} cx^2y^2/2|_{x^2}^{1}\, dx \\
&= \int_{-1}^{1} \frac{c}{2}x^2(1-x^4)\, dx = \frac{c}{2}[\frac{x^3}{3} - \frac{x^7}{7}]_{-1}^{1} = \frac{c}{2}[\frac{1}{3} - \frac{1}{7} - (-\frac{1}{3} + \frac{1}{7})] \\
&= c\frac{4}{21}.
\end{aligned}$$

Hence $c = 21/4$. The probability $P(X \geq Y)$ can be computed as

$$\begin{aligned}
P(X \geq Y) &= \int\int 1(x \geq y) \cdot 1(x^2 \leq y \leq 1)\frac{21}{4}x^2y\, dx\, dy \\
&= \int\int 1(x^2 \leq y \leq x \leq 1)\frac{21}{4}x^2y\, dx\, dy \\
&= \frac{21}{4}\int_0^1\int_{x^2}^{x} x^2y\, dy\, dx = \frac{21}{4}\int_0^1 x^2\frac{y^2}{2}\Big|_{x^2}^{x}\, dx \\
&= \frac{21}{8}\int_0^1 x^2(x^2 - x^4)\, dx = \frac{21}{8}\Big[\frac{x^5}{5} - \frac{x^7}{7}\Big]_0^1 \\
&= \frac{21}{8}\Big[\frac{1}{5} - \frac{1}{7}\Big] = \frac{3}{20} = 0.15.
\end{aligned}$$

Definition 26. The joint (cumulative) distribution function of X and Y is

$$\mathrm{cdf}_{X,Y}(x,y) = P(X \leq x, Y \leq y).$$

Example 43. Consider a bivariate function for $x = 1, 2, 3, 0 < y < 1$

$$f(x,y) = \frac{1}{3}xy^{x-1}.$$

Then the sum and integration of the function becomes

$$\sum_{x=1}^{3}\int_0^1 f(x,y)\, dy = \sum_{x=1}^{3}\int_0^1 \frac{1}{3}xy^{x-1}\, dy = \sum_{x=1}^{3}\frac{1}{3}y^x\Big|_0^1 = \sum_{x=1}^{3}\frac{1}{3} = 1.$$

Hence f is a joint pmf/pdf containing discrete X and continuous Y.

The probability of $X \geq 2, Y \geq 1/2$ is

$$\begin{aligned}
P(X \geq 2, Y \geq 1/2) &= \sum_{x=2}^{3} \int_{1/2}^{1} \frac{1}{3} x y^{x-1}\, dy = \sum_{x=2}^{3} \frac{1}{3} y^x \Big|_{1/2}^{1} \\
&= \sum_{x=2}^{3} \frac{1}{3}\Big(1 - (\frac{1}{2})^x\Big) = \frac{1}{3}\Big[1 - (\frac{1}{2})^2 + 1 - (\frac{1}{2})^3\Big] \\
&= \frac{13}{24} = 0.5417.
\end{aligned}$$

Theorem 22. Consider two random variables X and Y.

$$\begin{aligned}
\lim_{y\to-\infty} \mathrm{cdf}_{X,Y}(x,y) = 0 \quad & \lim_{y\to\infty} \mathrm{cdf}_{X,Y}(x,y) = \mathrm{cdf}_X(x) \\
\lim_{x\to-\infty} \mathrm{cdf}_{X,Y}(x,y) = 0 \quad & \lim_{x\to\infty} \mathrm{cdf}_{X,Y}(x,y) = \mathrm{cdf}_Y(y)
\end{aligned}$$

Proof. Note that $\mathrm{cdf}_{X,Y}(x,y) = P(X \leq x, Y \leq y)$. Then

$$\begin{aligned}
\lim_{y\to-\infty} \mathrm{cdf}_{X,Y}(x,y) &= \lim_{y\to-\infty} P(X \leq x, Y \leq y) \leq \lim_{y\to-\infty} P(Y \leq y) = 0. \\
\lim_{y\to\infty} \mathrm{cdf}_{X,Y}(x,y) &= \lim_{y\to\infty} P(X \leq x, Y \leq y) = P(X \leq x) = \mathrm{cdf}_X(x).
\end{aligned}$$

□

Example 44. Two random variables X and Y are defined on $[0,2]$ and the joint cdf is defined as

$$\mathrm{cdf}_{X,Y}(x,y) = \frac{1}{16} xy(x+y).$$

By sending x and y to infinite separately, $\mathrm{cdf}_X(x) = \lim_{y\to\infty} \mathrm{cdf}_{X,Y}(x,y) = \mathrm{cdf}_{X,Y}(x,2) = \frac{1}{8}x(x+2)$ and $\mathrm{cdf}_Y(y) = \lim_{x\to\infty} \mathrm{cdf}_{X,Y}(x,y) = \mathrm{cdf}_{X,Y}(2,y) = \frac{1}{8}y(y+2)$.

The joint density is

$$\begin{aligned}
\mathrm{pdf}_{X,Y}(x,y) &= \frac{\partial^2}{\partial x \partial y} \mathrm{cdf}_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} \frac{1}{16} xy(x+y) 1(0 \leq x, y \leq 2) \\
&= \frac{1}{8}(x+y) 1(0 \leq x, y \leq 2).
\end{aligned}$$

The densities of X and Y are

$$\begin{aligned}
\mathrm{pdf}_X(x) &= \frac{\partial}{\partial x} \mathrm{cdf}_X(x) = \frac{\partial}{\partial x} \frac{1}{8} x(x+2) = \frac{1}{4}(x+1) \\
\mathrm{pdf}_Y(y) &= \frac{\partial}{\partial y} \mathrm{cdf}_Y(y) = \frac{\partial}{\partial y} \frac{1}{8} y(y+2) = \frac{1}{4}(y+1)
\end{aligned}$$

Then

$$\mathrm{pdf}_X(x)\mathrm{pdf}_Y(y) = \frac{1}{16}(x+1)(y+1) \neq \frac{1}{8}(x+y) = \mathrm{pdf}_{X,Y}(x,y)$$

implies X and Y are not independent.

Marginal Distributions

Definition 27. Suppose X and Y are random variables. The cdf or pmf or pdf of X (or Y) derived from the joint cdf or pmf or pdf is called the *marginal* cdf or pmf or pdf of X (or Y) respectively.

Theorem 23. $\mathrm{cdf}_X(x) = \lim_{y\to\infty} \mathrm{cdf}_{X,Y}(x,y)$, $\mathrm{pmf}_X(x) = \sum_y \mathrm{pmf}_{X,Y}(x,y)$, $\mathrm{pdf}_X(x) = \int \mathrm{pdf}_{X,Y}(x,y)\, dy$

Proof. When (X,Y) is discrete, $\mathrm{pmf}_X(x) = P(X = x) = P(X = x, Y < \infty) = \sum_y P(X = x, Y = y) = \sum_y \mathrm{pmf}_{X,Y}(x,y)$. Similarly when (X,Y) is continuous, for any Borel set B

$$P(X \in B) = P(X \in B, Y < \infty) = \int_B \int_{-\infty}^{\infty} \mathrm{pdf}_{X,Y}(x,y)\, dy\, dx.$$

Hence $\mathrm{pdf}_X(x) = \int_{-\infty}^{\infty} \mathrm{pdf}_{X,Y}(x,y)\, dy$. Or the derivative of marginal cdf gives

$$\begin{aligned}
\mathrm{pdf}_X(x) &= \frac{\partial}{\partial x}\mathrm{cdf}_X(x) = \frac{\partial}{\partial x}\lim_{y\to\infty} \mathrm{cdf}_{X,Y}(x,y) = \lim_{y\to\infty}\frac{\partial}{\partial x}\mathrm{cdf}_{X,Y}(x,y) \\
&= \lim_{y\to\infty}\frac{\partial}{\partial x}\int_{-\infty}^{y}\int_{-\infty}^{x} \mathrm{pdf}_{X,Y}(u,v)\, du\, dv = \int_{-\infty}^{\infty} \mathrm{pdf}_{X,Y}(x,v)\, dv \\
&= \int_{-\infty}^{\infty} \mathrm{pdf}_{X,Y}(x,y)\, dy.
\end{aligned}$$

□

Definition 28. Two random variables X and Y are independent if and only if

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

for any Borel sets A and B.

Theorem 24. If two random variables X and Y are independent, then the follows hold if the functions exist.
(a) $\mathrm{cdf}_{X,Y}(x,y) = \mathrm{cdf}_X(x) \times \mathrm{cdf}_Y(y)$ for all x, y.
(b) $\mathrm{pmf}_{X,Y}(x,y) = \mathrm{pmf}_X(x) \times \mathrm{pmf}_Y(y)$ for all x, y.
(c) $\mathrm{pdf}_{X,Y}(x,y) = \mathrm{pdf}_X(x) \times \mathrm{pdf}_Y(y)$ for all x, y.

Proof. The definition of cdf implies

$$\begin{aligned}
\mathrm{cdf}_{X,Y}(x,y) &= P(X \le x, Y \le y) = P(X \in (-\infty, x], Y \in (-\infty, y]) \\
&= P(X \in (-\infty, x])P(Y \in (-\infty, y]) = \mathrm{cdf}_X(x)\mathrm{cdf}_Y(y).
\end{aligned}$$

Similarly, the joint probability mass function becomes

$$\begin{aligned}
\mathrm{pmf}_{X,Y}(x,y) &= P(X = x, Y = y) = P(X \in \{x\}, Y \in \{y\}) \\
&= P(X \in \{x\})P(Y \in \{y\}) = \mathrm{pmf}_X(x)\mathrm{pmf}_Y(y).
\end{aligned}$$

The pdf can be obtained by differentiation of cdf, that is,

$$\begin{aligned}\mathrm{pdf}_{X,Y}(x,y) &= \frac{\partial^2}{\partial x \partial y}\mathrm{cdf}_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y}\mathrm{cdf}_X(x)\mathrm{cdf}_Y(y)\\ &= \frac{\partial}{\partial x}\mathrm{cdf}_X(x)\frac{\partial}{\partial y}\mathrm{cdf}_Y(y) = \mathrm{pdf}_X(x)\mathrm{pdf}_Y(y).\end{aligned}$$

□

Theorem 25. Two random variables X and Y are independent if one of the follows hold.
(a) $\mathrm{cdf}_{X,Y}(x,y) = \mathrm{cdf}_X(x) \times \mathrm{cdf}_Y(y)$ for all x, y.
(b) $\mathrm{pmf}_{X,Y}(x,y) = \mathrm{pmf}_X(x) \times \mathrm{pmf}_Y(y)$ for all x, y.
(c) $\mathrm{pdf}_{X,Y}(x,y) = \mathrm{pdf}_X(x) \times \mathrm{pdf}_Y(y)$ for all x, y.

Proof. We already proved that (a) $\iff$ (b) or (c) and will only prove (c) implies independence.

Assume (c). For any Borel sets A and B,

$$\begin{aligned}P(X \in A, Y \in B) &= \int_A \int_B \mathrm{pdf}_{X,Y}(x,y)\; dy\; dx\\ &= \int_A \int_B \mathrm{pdf}_X(x)\mathrm{pdf}_Y(y)\; dy\; dx\\ &= \int_A \mathrm{pdf}_X(x)\; dx \int_B \mathrm{pdf}_Y(y)\; dy\\ &= P(X \in A)P(Y \in B).\end{aligned}$$

Hence X and Y are independent. □

Example 45. Two independent random variables X and Y have the same density $2x1(0 \le x \le 1)$. Compute $P(X+Y \le 1)$.

Note that $\{(x,y) : x+y \le 1, 0 \le x, y \le 1\} = \{(x,y) : 0 \le x \le 1, 0 \le y \le 1-x\}$. Hence

$$\begin{aligned}P(X+Y \le 1) &= \int_0^1 \int_0^{1-x} 2x \cdot 2y\; dy\; dx = \int_0^1 2xy^2\Big|_0^{1-x}\; dx\\ &= \int_0^1 2x[(1-x)^2 - 0^2]\; dx = 2\int_0^1 x - 2x^2 + x^3\; dx\\ &= 2[x^2/2 - 2x^3/3 + x^4/4]_0^1 = 2[1/2 - 2/3 + 1/4]\\ &= 1/6.\end{aligned}$$

Example 46. Two random variables X and Y have a joint pdf given by

$$\mathrm{pdf}_{X,Y}(x,y) = kx^2y^2 1(x^2+y^2 \le 1).$$

Are X and Y independent?

Easy answer: $\mathrm{pdf}_{X,Y}(0.7, 0.8) = 0$ because $0.7^2 + 0.8^2 = 1.13 > 1$. While $\mathrm{pdf}_Y(0.8) = \int_{-0.6}^{0.6} \mathrm{pdf}_{X,Y}(x, 0.8)\; dx = 1.28k\int_0^{0.6} x^2\; dx = 0.09216k > 0$ and $\mathrm{pdf}_X(0.7) = 0.1189k > 0$. Hence X and Y are not independent.

Even though X and Y are exchanged, the joint density doesn't change. Hence both X and Y have the same distribution.

The marginal density of X is

$$\begin{aligned}\mathrm{pdf}_X(x) &= \int_{-1}^{1} \mathrm{pdf}_{X,Y}(x,y)\ dy = \int_{-1}^{1} kx^2y^2 1(x^2+y^2 \le 1)\ dy \\ &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} kx^2y^2\ dy = kx^2y^3/3\Big|_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} = 2kx^2(1-x^2)^{3/2}/3.\end{aligned}$$

Note that

$$\begin{aligned}\mathrm{pdf}_X(x)\mathrm{pdf}_Y(y) &= 2kx^2(1-x^2)^{3/2}/3 \times 2ky^2(1-y^2)^{3/2}/3 \\ &\neq kx^2y^2 1(x^2+y^2 \le 1)\end{aligned}$$

Therefore X and Y are not independent.

Or the joint density $\mathrm{pdf}_{X,Y}(x,y) = kx^2y^2 1(x^2+y^2 \le 1)$ cannot decompose into functions of x and y separately. Thus X and Y are not independent.

Example 47. Two random variables X and Y have a joint density of the form

$$\mathrm{pdf}_{X,Y}(x,y) = ke^{-(x+2y)}1(x \ge 0, y \ge 0).$$

Determine the coefficient k. Determine whether or not X and Y are independent.

The probability of whole sample space gives

$$\begin{aligned}1 &= \int\int \mathrm{pdf}_{X,Y}(x,y)\ dx\ dy = \int_0^{\infty}\int_0^{\infty} ke^{-(x+2y)}\ dx\ dy \\ &= \int_0^{\infty} -ke^{-(x+2y)}\Big|_0^{\infty}\ dy = \int_0^{\infty} ke^{-2y}\ dy = -ke^{-2y}/2\Big|_0^{\infty} = k/2.\end{aligned}$$

Hence $k = 2$. The marginal density of Y is

$$\begin{aligned}\mathrm{pdf}_Y(y) &= \int \mathrm{pdf}_{X,Y}(x,y)\ dx = \int_0^{\infty} 2e^{-(x+2y)}\ dx = -2e^{-(x+2y)}\Big|_0^{\infty} \\ &= 2e^{-2y}.\end{aligned}$$

Similarly the marginal density of X is

$$\begin{aligned}\mathrm{pdf}_X(x) &= \int \mathrm{pdf}_{X,Y}(x,y)\ dy = \int_0^{\infty} 2e^{-(x+2y)}\ dy = -e^{-(x+2y)}\Big|_0^{\infty} \\ &= e^{-x}.\end{aligned}$$

Conditional Distributions

Two discrete random variables X and Y have a joint probability mass function $f(x,y)$. The conditional probability of x given $Y = y$ is given by

$$P(X = x \,|\, Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{\mathrm{pmf}_{X,Y}(x,y)}{\mathrm{pmf}_Y(y)}.$$

Hence the conditional probability mass function becomes

$$\mathrm{pmf}_{X\,|\,Y}(x\,|\,y) = P(X = x\,|\,Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{\mathrm{pmf}_{X,Y}(x, y)}{\mathrm{pmf}_Y(y)}.$$

Continuous case is a bit complicated due to $P(Y = y) = 0$ in general.

Two random variables X and Y have a joint density $f(x, y)$. The conditional probability of x given $Y = y$ wasn't be able to be defined previously because $P(Y = y) = 0$. Consider $B_\Delta = (y - \Delta, y + \Delta)$. If the density $\mathrm{pdf}_Y(y)$ is continuous and positive at y, then $P(Y \in B_\Delta) > 0$.

Considering $P(Y \in B_\Delta) \approx \mathrm{pdf}_Y(y)2\Delta$ for very small $\Delta > 0$. Then the conditional probability of $X \in A$ is

$$\begin{aligned}
P(X \in A\,|\,Y \in B_\Delta) &= \frac{P(X \in A, Y \in B_\Delta)}{P(Y \in B_\Delta)} = \frac{\int_{B_\Delta}\int_A \mathrm{pdf}_{X,Y}(x, y)\;dx\;dy}{\int_{B_\Delta}\mathrm{pdf}_Y(y)\;dy} \\
&\approx \frac{\int_A \mathrm{pdf}_{X,Y}(x, y)\;dx \times 2\Delta}{\mathrm{pdf}_Y(y) \times 2\Delta} = \int_A \frac{\mathrm{pdf}_{X,Y}(x, y)}{\mathrm{pdf}_Y(y)}\;dx
\end{aligned}$$

Definition 29. The condition conditional density of X given $Y = y$ is

$$\mathrm{pdf}_{X\,|\,Y}(x\,|\,y) = \frac{\mathrm{pdf}_{X,Y}(x, y)}{\mathrm{pdf}_Y(y)}.$$

The definition of conditional density implies the following theorem.

Theorem 26. $\mathrm{pdf}_{X,Y}(x, y) = \mathrm{pdf}_X(x)\mathrm{pdf}_{Y\,|\,X}(y\,|\,x)$.

Example 48. In a coffee shop, there is only one server working. The waiting time of each client is denoted by X which is depending on the working rate of the server denoted by Y. Suppose the density of X is $y\exp(-yx)1(x > 0)$ given $Y = y$. What is the joint density of X and Y if Y has density $e^{-y}1(y > 0)$?

The joint density is given by

$$\mathrm{pdf}_{X,Y}(x, y) = \mathrm{pdf}_Y(y)\mathrm{pdf}_{X\,|\,Y}(x\,|\,y) = e^{-y}1(y > 0) \times ye^{-yx}1(x > 0).$$

Multivariate Distributions

Definition 30. The joint cumulative distribution function of n random variables $X_1, \ldots, X_n$ is defined by

$$\mathrm{cdf}_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 \le x_1, \ldots, X_n \le x_n).$$

The joint probability mass/density function of n discrete/continuous random variables $X_1, \ldots, X_n$ is

$$\begin{aligned}
&\mathrm{pmf}_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n) \\
&P((X_1, \ldots, X_n) \in B) = \underset{B}{\int \cdots \int} \mathrm{pdf}_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\;dx_n\;\cdots\;dx_1.
\end{aligned}$$

Definition 31. Let $X_1, \ldots, X_n$ be random variables. Marginal cumulative distribution, probability mass, probability density functions of $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$ are

$$\begin{aligned}
&\mathrm{cdf}_{X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n}(x_1, \ldots, x_i - 1, x_{i+1}, \ldots, x_n) \\
&= \lim_{x_i \to \infty} \mathrm{cdf}_{X_1,\ldots,X_{i-1},X_i,X_{i+1},\ldots,X_n}(x_1, \ldots, x_i - 1, x_i, x_{i+1}, \ldots, x_n), \\
&\mathrm{pmf}_{X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n}(x_1, \ldots, x_i - 1, x_{i+1}, \ldots, x_n) \\
&= \sum_{x_i} \mathrm{pmf}_{X_1,\ldots,X_{i-1},X_i,X_{i+1},\ldots,X_n}(x_1, \ldots, x_i - 1, x_i, x_{i+1}, \ldots, x_n), \\
&\mathrm{pdf}_{X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n}(x_1, \ldots, x_i - 1, x_{i+1}, \ldots, x_n) \\
&= \int \mathrm{pmf}_{X_1,\ldots,X_{i-1},X_i,X_{i+1},\ldots,X_n}(x_1, \ldots, x_i - 1, x_i, x_{i+1}, \ldots, x_n)\, dx_i.
\end{aligned}$$

Theorem 27. Let $X_1, \ldots, X_n$ be continuous random variables having cdf. Then

$$\mathrm{pdf}_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(x_1, \ldots, x_n)$$

The second fundamental theorem of calculus guarantees the theorem.

Example 49. Suppose there is a machine having three parts, say $i = 1, 2, 3$. Let X_i be the time to fail working for $i = 1, 2, 3$. Assume the joint cdf is $\mathrm{cdf}_{X_1,X_2,X_3}(x_1, x_2, x_3) = (1 - e^{-x_1})(1 - e^{-2x_2})(1 - e^{-3x_3})1(x_1, x_2, x_3 \geq 0)$.

The density is $\mathrm{pdf}_{X_1,X_2,X_3}(x_1, x_2, x_3) = \frac{\partial^3}{\partial x_1 \partial x_2 \partial x_3} \mathrm{cdf}_{X_1,X_2,X_3}(x_1, x_2, x_3) = 6e^{-x_1 - 2x_2 - 3x_3} 1(x_1, x_2, x_3 \geq 0)$.

The marginal cdf of X_1 is $\mathrm{cdf}_{X_1}(x_1) = \lim_{x_2, x_3 \to \infty} \mathrm{cdf}_{X_1,X_2,X_3}(x_1, x_2, x_3) = 1 - e^{-x_1}$ and the marginal pdf of X_1 is

$$\mathrm{pdf}_{X_1}(x_1) = \frac{\partial}{\partial x_1} \mathrm{cdf}_{X_1}(x_1) = \frac{d}{dx_1}(1 - e^{-x_1}) = e^{-x_1}.$$

Or by the definition

$$\mathrm{pdf}_{X_1}(x_1) = \int_0^\infty \int_0^\infty 6e^{-x_1 - 2x_2 - 3x_3}\, dx_2\, dx_3 = e^{-x_1}.$$

Definition 32. Random variables $X_1, \ldots, X_n$ are *independent* if and only if for any Borel sets $B_1, \ldots, B_n$

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = P(X_1 \in B_1) \cdots P(X_n \in B_n)$$

Theorem 28. Random variables $X_1, \ldots, X_n$ are *independent* if and only if

$$\mathrm{cdf}_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \mathrm{cdf}_{X_1}(x_1) \cdots \mathrm{cdf}_{X_n}(x_n).$$

Functions of Random Variable

Rather than all outcomes, some random variables are of our interest. Furthermore, some transformed random variables are much more important in many problems.

Example 50. Let $X \sim \text{uniform}(\{1, 2, \ldots, 9\})$. The distance from the middle is our interest, that is, $Y = |X - 5|$. For example the probability of $Y \leq 1$ is $P(Y \leq 1) = P(|X - 5| \leq 1) = P(X \in \{4, 5, 6\}) = 3/9 = 1/3$.

Theorem 29. Let X be a discrete random variable and $Y = g(X)$ be a transformed random variable where $g : \mathbb{R} \to \mathbb{R}$ is a function. The probability mass function of Y is

$$\text{pmf}_Y(y) = \sum_{x:g(x)=y} \text{pmf}_X(x).$$

Example 51. $\text{pmf}_Y(0) = P(Y = 0) = P(X = 5) = 1/9$ and $\text{pmf}_Y(1) = P(Y = 1) = P(X \in \{4, 6\}) = 2/9$. Hence $P(Y \leq 1) = P(Y \in \{0, 1\}) = 1/9 + 2/9 = 1/3$.

Similarly new random variables can be derived from a continuous random variable. Then the density and distribution functions can be defined in a similar fashion.

Theorem 30. Let X be a continuous random variable and $Y = g(X)$ be a transformed continuous random variable where g is an appropriate transformation like continuous increasing. The cumulative distribution function of Y is

$$\text{cdf}_Y(y) = \int_{\{x:g(x) \leq y\}} \text{pdf}_X(x)\, dx.$$

The probability density function of Y is

$$\text{pdf}_Y(y) = \frac{d}{dy} \text{cdf}_Y(y).$$

Example 52. Let X be a continuous random variable with $\text{pdf}_X(x) = e^{-x} 1(x > 0)$. Then cumulative distribution function of X is

$$\text{cdf}_X(x) = \int_0^x e^{-z}\, dz = -e^{-z}\Big|_0^x = -e^{-x} + 1.$$

For $g(x) = 1 - e^{-x}$, define $Y = g(X)$. Then, for $y \in (0, 1)$, $\text{cdf}_Y(y) = P(Y \leq y) = P(X \leq -\log(1 - y)) = 1 - e^{-(-\log(1-y))} = 1 - (1 - y) = y$, that is, $Y \sim \text{uniform}(0, 1)$. So $\text{pdf}_Y(y) = \frac{d}{dy}\text{cdf}_Y(y) = 1(0 < y < 1)$.

Exercise 14. Let X be a continuous random variable. Show that $\text{cdf}_X(x)$ is also continuous.

Theorem 31. Let X be a continuous random variable and $F(x) = \text{cdf}_X(x)$. Then new random variable $Y = F^{-1}(X)$ is uniformly distributed on $(0, 1)$, that is, $Y \sim \text{uniform}(0, 1)$.

Proof. Note F is continuous and non-decreasing function, that is, there exists F^{-1} which is also non-decreasing. For any $y \in (0, 1)$, there exists x_0 such that $F(x_0) = y$. It is easy to check that $\{Y \leq y\} = \{X \leq x_0\}$. Then

$$\text{cdf}_Y(y) = P(Y \leq y) = P(X \leq x_0) = F(x_0) = y.$$

Which implies $\text{pdf}_Y(y) = \frac{d}{dy}\text{cdf}_Y(y) = 1(0 < y < 1)$. □

Exercise 15. Let X be a random variable. Define $F(x) = \text{cdf}_X(x)$ and $G(y) = \inf\{x : F(x) \geq y\}$ for any $y \in (0, 1)$. If $Y \sim \text{uniform}(0, 1)$, then the distribution function of $Y = G(Y)$ is F, that is, $\text{cdf}_Y(y) = F(y)$.

Theorem 32 (change of variable)**.** Let X be a continuous random variable and g be a one-to-one and differentiable function. Then the density of random variable $Y = g(X)$ is

$$\text{pdf}_Y(y) = \text{pdf}_X(g^{-1}(y))\Big|\frac{d}{dy}g^{-1}(y)\Big|$$

whenever y is in the range of $Y(S)$

Proof. The function g is either increasing or decreasing. Assume g is increasing. Note that $\text{cdf}_Y(y) = P(Y \leq y) = P(X \leq g^{-1}(y)) = \text{cdf}_X(g^{-1}(y))$ and

$$\text{pdf}_Y(y) = \frac{d}{dy}\text{cdf}_X(g^{-1}(y)) = \text{pdf}_X(g^{-1}(y))\Big|\frac{d}{dy}g^{-1}(y)\Big|.$$

If g is decreasing, the $\text{cdf}_Y(y) = P(X \geq g^{-1}(y)) = 1 - \text{cdf}_X(g^{-1}(y))$ and

$$\text{pdf}_Y(y) = \frac{d}{dy}\text{cdf}_X(g^{-1}(y)) = \text{pdf}_X(g^{-1}(y))\Big|\frac{d}{dy}g^{-1}(y)\Big|.$$

$\square$

Example 53 (Microbial Growth)**.** At time 0, suppose that v organisms are introduced into a large tank of water, and let X be the rate of growth. After time t, we would predict a population size of ve^{Xt}. Assume that X is unknown but has a continuous distribution with pdf

$$\text{pdf}_X(x) = 3(1-x)^2 1(0 < x < 1).$$

We are interested in the distribution of $Y = ve^{Xt}$ for known values of v and t. Then $y = ve^{tx}$ solves $x = \log(y/v)/t$ and the change of variable formula gives

$$\begin{aligned}\text{pdf}_Y(y) &= 3(1 - \frac{\log(y/v)}{t})^2 1(0 < \frac{\log(y/v)}{t} < 1)\Big|\frac{d}{dy}\frac{\log(y/v)}{t}\Big| \\ &= 3(1 - \frac{\log(y/v)}{t})^2 1(1 < y < ve^t)\frac{1}{yt}.\end{aligned}$$

Basically change of variable can be written as, for $y = g(x)$,

$$\text{pdf}_Y(y) = \text{pdf}_X(x)\Big|\frac{dx}{dy}\Big|.$$

Even for multivariate random vector $\mathbf{y} = g(\mathbf{x})$,

$$\text{pdf}_{\mathbf{Y}}(\mathbf{y}) = \text{pdf}_{\mathbf{X}}(\mathbf{x})\Big|\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\Big|.$$

Then

$$\int \text{pdf}_{\mathbf{Y}}(\mathbf{y})\, d\mathbf{y} = \int \text{pdf}_{\mathbf{X}}(\mathbf{x})\Big|\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\Big|\, d\mathbf{y} = \int \text{pdf}_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} = 1.$$

Functions of Two or More Random Variables

From a set of data $X_1, \dots, X_n$, a few transformations are often of interest.

Theorem 33. Consider discrete random variables $X_1, \dots, X_n$. There exist m functions $g_1, \dots, g_m$ so that $Y_i = g_i(X_1, \dots, X_n)$. The joint probability mass function of $\mathbf{Y} = (Y_1, \dots, Y_m)$ is

$$\mathrm{pmf}_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{x}: g_i(\mathbf{x}) = y_i, i=1,\dots,m} \mathrm{pmf}_{\mathbf{X}}(\mathbf{x}).$$

Definition 33. Random variables $X_1, \dots, X_n$ are said to be *independent and identically distributed* (i.i.d.) if all random variables have the same distribution and are independent.

Example 54. Suppose $X_i \sim i.i.d.$ Bernoulli(p). The sum $Y_n = X_1 + \cdots + X_n$ satisfies $Y \sim$ binomial(n, p).

Obviously $Y_1 = X_1 \sim$ Bernoulli(p) $\sim$ binomial($1, p$). Assume $Y_k \sim$ binomial(k, p). Then

$$\begin{aligned} P(Y_{k+1} = 0) &= P(Y_k = 0, X_{k+1} = 0) = P(Y_k = 0)P(X_{k+1} = 0) \\ &= \binom{k}{0} p^0 (1-p)^k \times (1-p) = \binom{k+1}{0} p^0 (1-p)^{k+1}. \end{aligned}$$

For $j = 1, \dots, k+1$,

$$\begin{aligned} P(Y_{k+1} = j) &= P(Y_k = j, X_{k+1} = 0 \text{ or } Y_k = j-1, X_{k+1} = 1) \\ &= P(Y_k = j, X_{k+1} = 0) + P(Y_k = j-1, X_{k+1} = 1) \\ &= P(Y_k = j)P(X_{k+1} = 0) + P(Y_k = j-1)P(X_{k+1} = 1) \\ &= \binom{k}{j} p^j (1-p)^{k-j} \cdot (1-p) + \binom{k}{j-1} p^{j-1} (1-p)^{k-(j-1)} \cdot p \\ &= (\binom{k}{j} + \binom{k}{j-1}) p^j (1-p)^{k+1-j} \\ &= \binom{k+1}{j} p^j (1-p)^{k+1-j} \end{aligned}$$

Hence $Y_{k+1} \sim$ binomial($k+1, p$).

Theorem 34. Let X and Y be two independent continuous random variables. The density of $Z = X + Y$ is

$$\mathrm{pdf}_Z(z) = \int \mathrm{pdf}_X(x) \mathrm{pdf}_Y(z-x)\, dx.$$

Proof. Consider a change of variable $(x, y) \mapsto (x, x+y) = (x, z)$, then the joint density becomes

$$\begin{aligned} \mathrm{pdf}_{X,Z}(x, z) &= \mathrm{pdf}_{X,Y}(x, z-x) \left| \frac{\partial(x,y)}{\partial(x,z)} \right| = \mathrm{pdf}_X(x) \mathrm{pdf}_Y(z-x) \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} \\ &= \mathrm{pdf}_X(x) \mathrm{pdf}_Y(z-x). \end{aligned}$$

The marginal density of Z is obtained by integrating x out, that is,

$$\text{pdf}_Z(z) = \int \text{pdf}_{X,Z}(x,z)\, dx = \int \text{pdf}_X(x)\text{pdf}_Y(z-x)\, dx.$$

$\square$

Example 55. Suppose two independent random variables X and Y having the same density

$$\text{pdf}_X(x) = \text{pdf}_Y(x) = e^{-x}1(x>0).$$

The density of $Z = X + Y$ is

$$\begin{aligned}\text{pdf}_Z(z) &= \int \text{pdf}_X(x)\text{pdf}_Y(z-x)\, dx \\ &= \int_{-\infty}^{\infty} e^{-x}1(x>0)e^{-(z-x)}1(z-x>0)\, dx = \int_0^z e^{-z}\, dx \\ &= ze^{-z}.\end{aligned}$$

Example 56. Let $X_1, \ldots, X_n$ be a i.i.d. with the common distribution function F. Define

$$Y_n = \max\{X_1,\ldots,X_n\} \text{ and } Y_1 = \min\{X_1,\ldots,X_n\}$$

Find the cumulative distribution functions of Y_1 and Y_n.

The definition implies

$$\begin{aligned}P(Y_n \le y) &= P(\max\{X_1,\ldots,X_n\} \le y) = P(X_1 \le y, \ldots, X_n \le y) \\ &= P(X_1 \le y)\cdots P(X_n \le y) = [\text{cdf}_X(y)]^n = [F(y)]^n.\end{aligned}$$

Similarly

$$\begin{aligned}P(Y_1 \le y) &= 1 - P(Y_1 > y) = 1 - P(\min\{X_1,\ldots,X_n\} > y) \\ &= 1 - P(X_1 > y, \ldots, X_n > y) \\ &= 1 - P(X_1 > y)\cdots P(X_n > y) = 1 - [1 - \text{cdf}_X(y)]^n \\ &= 1 - [1 - F(y)]^n.\end{aligned}$$

The joint cumulative distribution function is

$$\begin{aligned}&\text{cdf}_{Y_1,Y_n}(y_1,y_2) = P(Y_1 \le y_1, Y_2 \le y_2) \\ &= P(Y_2 \le y_2) - P(Y_1 > y_1, Y_2 \le y_2) \\ &= [F(y_2)]^n - P(\min\{X_1,\ldots,X_n\} > y_1, \max\{X_1,\ldots,X_n\} \le y_2) \\ &= [F(y_2)]^n - P(y_1 < X_i \le y_2, i = 1, \ldots, n) \\ &= [F(y_2)]^n - [F(y_2) - F(y_1)]^n.\end{aligned}$$

If there exists a density, say $f(x) = \frac{d}{dx}F(x)$, then

$$\begin{aligned}
\mathrm{pdf}_{Y_1,Y_n}(y_1,y_2) &= \frac{\partial^2}{\partial y_1 \partial y_2}\mathrm{cdf}_{Y_1,Y_n}(y_1,y_2) \\
&= n(n-1)f(y_1)(F(y_2)-F(y_1))^{n-2}f(y_2).
\end{aligned}$$

Example 57. Suppose $X_1,\dots,X_n$ is an i.i.d. sample having density $\mathrm{pdf}_X(x) = e^{-x}1(x>0)$. The joint density of $Y_1 = \min(X_1,\dots,X_n)$ and $Y_n = \max(X_1,\dots,X_n)$ is

$$\begin{aligned}
\mathrm{pdf}_{Y_1,Y_n}(y_1,y_2) &= n(n-1)f(y_1)(F(y_2)-F(y_1))^{n-2}f(y_2) \\
&= n(n-1)e^{-y_1}(e^{-y_1}-e^{-y_2})^{n-2}e^{-y_2}1(0<y_1\le y_n)
\end{aligned}$$

and the marginal densities are

$$\begin{aligned}
\mathrm{pdf}_{Y_1}(y) &= nf(y)(1-F(y))^{n-1} = ne^{-y}e^{-(n-1)y} = ne^{-ny}, \\
\mathrm{pdf}_{Y_n}(y) &= nf(y)(F(y))^{n-1} = ne^{-y}(1-e^{-y})^{n-1}.
\end{aligned}$$

Theorem 35 (change of variable)**.** Suppose $X_1,\dots,X_n$ have a joint density function $f(x_1,\dots,x_n)$ and $Y_i = g_i(X_1,\dots,X_n)$ for one-to-one correspondent and differentiable functions g_i's, say $\mathbf{y} = g(\mathbf{x})$. The joint density of $Y_1,\dots,Y_n$ is

$$\mathrm{pdf}_{\mathbf{Y}}(\mathbf{y}) = \mathrm{pdf}_{\mathbf{X}}(\mathbf{x})\left|\det\Big(\frac{\partial(x_1,\dots,x_n)}{\partial(y_1,\dots,y_n)}\Big)\right|$$

where $\mathbf{x} = (x_1,\dots,x_n) = g^{-1}(\mathbf{y})$.

Example 58. Two continuous random variables X_1, X_2 have joint density $\mathrm{pdf}_{X_1,X_2}(x_1,x_2) = 4x_1x_2 1(0 < x_1, x_2 < 10$. Find the joint density of $Y_1 = X_1/X_2$ and $Y_2 = X_1X_2$.

Note that $x_1 = (y_1y_2)^{1/2}$ and $x_2 = (y_2/y_1)^{1/2}$.

$$\begin{aligned}
\mathrm{pdf}_{Y_1,Y_2}(y_1,y_2) &= \mathrm{pdf}_{X_1,X_2}((y_1y_2)^{1/2},(y_2/y_1)^{1/2})\begin{vmatrix} (y_2/y_1)^{1/2}/2 & (y_1/y_2)^{1/2}/2 \\ -(y_2/y_1^3)^{1/2}/2 & 1/[2(y_1y_2)^{1/2}] \end{vmatrix} \\
&= 4y_2 1(0<y_1<y_2<1)\frac{1}{2y_1} = 1(0<y_1<y_2<1)2y_2/y_1.
\end{aligned}$$

Example 59. Let $X_1,\dots,X_n$ be a random sample having density f. The order statistics $Y_1 \le Y_2 \le \cdots \le Y_n$ of $X_1,\dots,X_n$ have joint density

$$\mathrm{pdf}_{\mathbf{Y}}(\mathbf{y}) = n!f(y_1)\cdots f(y_n)1(y_1\le y_2\le\cdots\le y_n).$$

The joint density of Y_i, Y_j

$$\begin{aligned}
\mathrm{pdf}_{Y_i,Y_j}(x,z) &= \frac{n!}{(i-1)!1!(j-i-1)!1!(n-j)!} \\
&\quad\times [F(x)]^{i-1}f(x)[F(z)-F(x)]^{j-i-1}f(z)[1-F(z)]^{n-j}
\end{aligned}$$

is obtained using distribution function.

Example 60. For $\mathbf{X} = (X_1, \ldots, X_n)$ define $\mathbf{Y} = A\mathbf{X}$ for an invertible $n \times n$ matrix A. Then $\mathbf{X} = A^{-1}\mathbf{Y}$ and

$$\mathrm{pdf}_{\mathbf{Y}}(\mathbf{y}) = \mathrm{pdf}_{\mathbf{X}}(A^{-1}\mathbf{y})|\det(A^{-1})| = \mathrm{pdf}_{\mathbf{X}}(A^{-1}\mathbf{y})/|\det(A)|.$$

Exercises. (DS) 3.1.2, 3.1.4, 3.1.9, 3.1.10, 3.1.11, 3.2.4, 3.2.8, 3.2.9, 3.2.10, 3.2.13, 3.3.5, 3.3.6, 3.3.14, 3.3.15, 3.3.17, 3.3.18, 3.4.2, 3.4.3, 3.4,4, 3.4.5, 3.4.7, 3.4.8, 3.4.10, 3.5.1, 3.5.2, 3.5.3, 3.5.4, 3.5.6, 3.5.8, 3.5.15, 3.6.3, 3.6.4, 3.6.6, 3.6.7, 3.6.11, 3.6.12, 3.7.3, 3.7.6, 3.7.7, 3.7.8, 3.7.11, 3.7.12, 3.8.6, 3.8.7, 3.8.8, 3.8.13, 3.8.15, 3.8.17, 3.9.6, 3.9.8, 3.9.12, 3.9.14, 3.9.15, 3.9.16, 3.9.21, 3.11.3, 3.11.4, 3.11.14, 3.11.16, 3.11.22, 3.11.25, 3.11.26; (GS) 2.1.4, 2.1.5, 2.3.1, 2.3.2, 2.3.3, 2.3.4, 2.3.5, 2.4.1, 2.4.2, 1.3.1, 1.3.5, 3.2.1, 3.2.3, 3.2.5, 2.5.2, 2.5.6, 2.7.4, 2.7.6, 2.7.9, 2.7.13, 2.7.19, 2.7.20, 3.5.2, 3.5.3, 3.6.3, 3.6.4, 3.6.5, 3.6.7, 3.6.8, 3.7.5, 3.8.3, 3.11.9, 3.11.10, 3.11.28, 4.1.4, 4.2.4, 4.4.1, 4.4.5, 4.5.3, 4.6.4, 4.6.5, 4.6.6, 4.6.9, 4.7.7, 4.7.9, 4.7.14, 4.8.3, 4.8.5, 4.10.6, 4.11.3, 4.11.4, 4.11.10, 4.13.2, 4.13.12, 4.14.3, 4.14.12, 4.14.21, 4.14.24; (Ri) 2.2, 2.4, 2.6, 2.8, 2.9, 2.11, 2.14, 2.15, 2.16, 2.18, 2.20, 2.23, 2.28, 2.31, 2.33, 2.34, 2.35, 2.36, 2.38; (RM) 2.5.4, 2.5.5, 2.5.6, 2.5.11, 2.5.12, 2.5.13, 2.5.15, 2.5.16, 2.5.17, 2.5.21, 2.5.24, 2.5.27, 2.5.28, 2.5.29, 2.5.32, 2.5.33, 2.5.36, 2.5.38, 2.5.39, 2.5.40, 2.5.44, 2.5.47, 2.5.49, 2.5.51, 2.5.54, 2.5.55, 2.5.58, 2.5.59, 2.5.60, 2.5.62, 2.5.64, 2.5.67, 2.5.68, 3.8.4, 3.8.5, 3.8.6, 3.8.7, 3.8.8, 3.8.11, 3.8.12, 3.8.15, 3.8.19, 3.8.22, 3.8.23, 3.8.25, 3.8.27, 3.8.34, 3.8.37, 3.8.40, 3.8.42, 3.8.43, 3.8.46, 3.8.47, 3.8.51, 3.8.59, 3.8.64, 3.8.69, 3.8.72, 3.8.74, 3.8.80; (T) 7.20, 7.21, 7.22, 7.26, 7.30, 7.40, 7.41, 7.45, 7.48, 7.52, 8.2, 8.5, 8.7, 8.9, 8.14, 8.22, 8.24, 8.27, 8.29, 8.32, 8.34, 8.38, 8.39, 9.3, 9.4, 9.36, 9.37, 9.45, 9.48, 10.2, 10.4, 10.7, 10.8, 10.31, 10.35, 10.36, 10.40, 10.44, 10.46, 11.3, 11.5, 11.7, 11.10, 11.11, 11.13, 11.15, 11.17, 11.18, 13.4, 13.6, 13.7, 13.13, 14.2, 14.4;