If X has a continuous distribution with density function $f(x)$, then $\boldsymbol{F(x) = \int_{-\infty}^{x} f(t)\,dt}$ and $F(x)$ is a continuous, differentiable, non-decreasing function such that $\boldsymbol{\frac{d}{dx}F(x) = F'(x) = -S'(x) = f(x)}$. If X has a mixed distribution, then $F(x)$ is continuous except at the points of non-zero probability mass, where $F(x)$ will have a jump.

For any cdf $\boldsymbol{P[a < X \le b] = F(b) - F(a)}$, $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$.

Examples of probability, density and distribution functions:

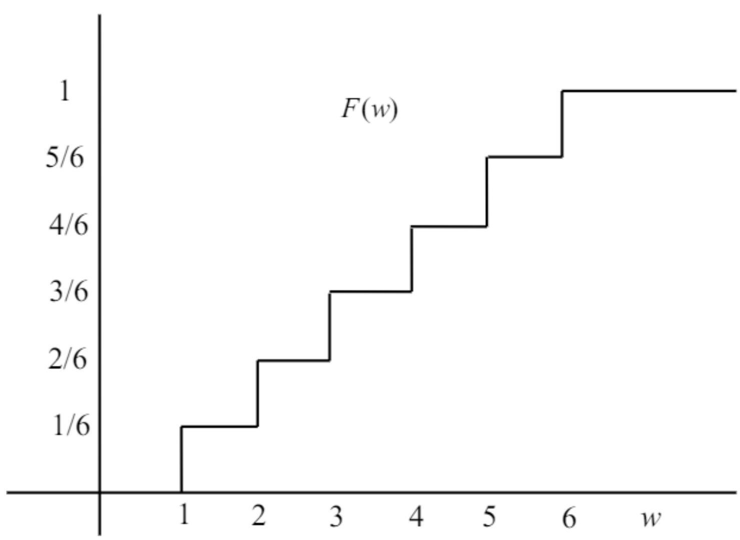
Example 4-3:

Discrete Random Variable on a Finite Number of Points (finite support)

W = number turning up when tossing one fair die, so W has probability function
$p(w) = p_W(w) = P[W = w] = \frac{1}{6}$ for $w = 1, 2, 3, 4, 5, 6$

$$F_W(w) = P[W \le w] = \begin{cases} 0 \ \text{ if } \ w < 1 \\ 1/6 \ \text{ if } \ 1 \le w < 2 \\ 2/6 \ \text{ if } \ 2 \le w < 3 \\ 3/6 \ \text{ if } \ 3 \le w < 4 \\ 4/6 \ \text{ if } \ 4 \le w < 5 \\ 5/6 \ \text{ if } \ 5 \le w < 6 \\ 1 \ \text{ if } \ w \ge 6 \end{cases}$$

The graph of the cdf (cumulative distribution function) is a step-function that increases at each point of probability by the amount of probability at that point (all 6 points have probability $\frac{1}{6}$ in this example). Since the support of W is finite (the "support" is the region of non-zero probability; for W that is the set of integers from 1 to 6), $F_W(w)$ reaches 1 at the largest point $W = 6$ (and stays at 1 for all $w \ge 6$).



□

Example 4-4:

Discrete Random Variable on an Infinite Number of Points (infinite support)

$X =$ number of successive independent tosses of a fair coin until the first head turns up.

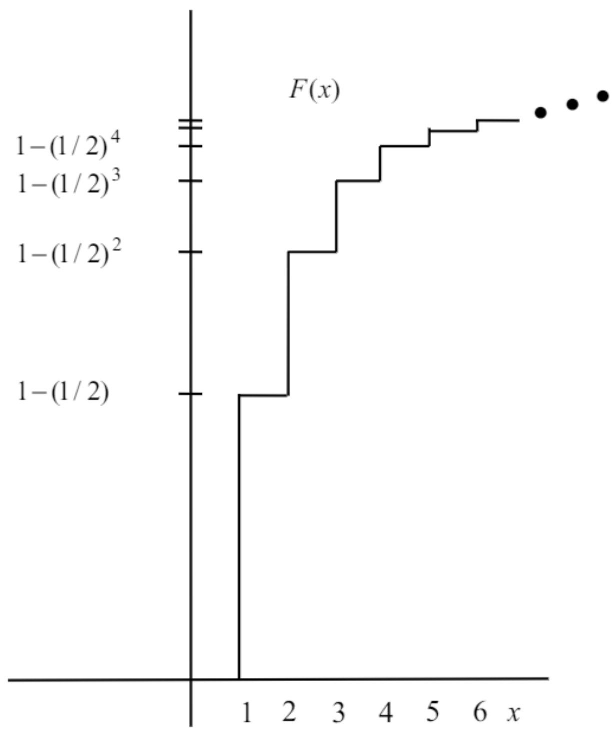
X can be any integer ≥ 1, and the probability function of X is $p_X(x) = \frac{1}{2^x}$, since

$P[\text{first head on toss } x] = P[(\text{toss } 1{,}T) \cap (\text{toss } 2{,}T) \cap \cdots \cap (\text{toss } x-1{,}T) \cap (\text{toss } x{,}H)]$

$= P[\text{toss } 1{,}T] \times P[\text{toss } 2{,}T] \times \cdots \times P[\text{toss } x-1{,}T] \times P[\text{toss } x{,}H] = \frac{1}{2} \times \frac{1}{2} \times \cdots \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{2^x}$.

The cdf is $F_X(x) = P[X \leq x] = P[X = 1] + P[X = 2] + \cdots + P[X = x] = \sum_{k=1}^{x} \frac{1}{2^k} = 1 - \frac{1}{2^x}$

for $x = 1, 2, 3, \ldots$. The graph of this cdf is a step-function that increases at each point of probability by the amount of probability at that point. Since the support of X is infinite (the support in this case is the set of integers ≥ 1) $F_X(x)$ never reaches 1, but approaches 1 as a limit as $x \to \infty$. The graph of $F_X(x)$ is below.



The probability that the first head occurs on an even numbered toss is

$$P(X \text{ is even}) = P(X = 2, 4, 6, \ldots) = P(X = 2) + P(X = 4) + P(X = 6) + \cdots$$
$$= \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \cdots = \frac{1}{2^2} \times [1 + \frac{1}{2^2} + (\frac{1}{2^2})^2 + \cdots] = \frac{1}{2^2} \Big/ [1 - \frac{1}{2^2}] = \frac{1}{3}$$

The probability that the first head occurs on, or after the k-th toss ($k = 1, 2, \ldots$) is

$$P(X \geq k) = P(X = k) + P(X = k+1) + P(X = k+2) + \cdots$$
$$= \frac{1}{2^k} + \frac{1}{2^{k+1}} + \frac{1}{2^{k+2}} + \cdots = \frac{1}{2^k} \times [1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots] = \frac{1}{2^k} \times 2 = \frac{1}{2^{k-1}}$$

for $k = 1, 2, 3, \ldots$ □

The typical behavior of the cdf $F(x)$ is to tend to increase toward 1 as x increases. Depending on the nature of the random variable, $F(x)$ may actually reach 1 at some point, as in Example 4-3, or $F(x)$ might approach 1 as a limit, as in Example 4-4. For a continuous random variable, $F(x)$ has similar increasing behavior, but will be increasing continuously rather than in the series of steps we have seen for a discrete random variable. $F(x)$ will never decrease, but it may remain "flat" for a while, as can be seen in the previous two examples.

Example 4-5:

Continuous Random Variable on a Finite Interval

Y is a continuous random variable on the interval $(0,1)$ with density function

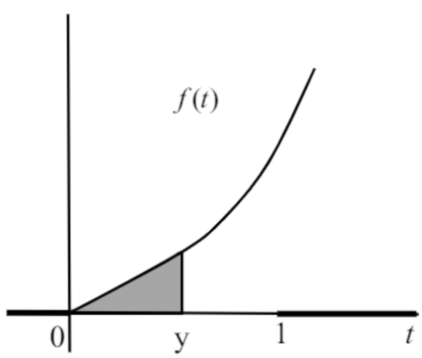
$$f_Y(y) = \begin{cases} 3y^2 \text{ for } 0<y<1 \\ 0, \text{ elsewhere} \end{cases}.$$

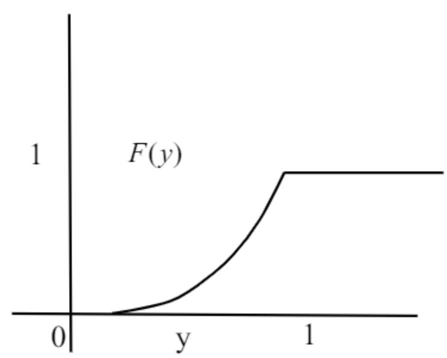
The cdf is $F_Y(y) = \int_{-\infty}^{y} f(t)\,dt = \int_0^y 3t^2\,dt = y^3$ if $y \leq 1$. Then

$$F_Y(y) = \int_0^y f_Y(t)\,dt = \begin{cases} 0 \text{ if } y < 0 \\ y^3 \text{ if } 0 \leq y < 1 \\ 1 \text{ if } y \geq 1 \end{cases}.$$

The graphs of $f_Y(y)$ and $F_Y(y)$ are as follows.

The heavy line in the graph of $f_Y(y)$ indicates that the density is 0 outside the interval $(0,1)$. Note that the cdf increases continuously, reaching 1 at the right end of the interval for the probability space.





Some other probabilities are $P(Y \leq \frac{1}{2}) = F(\frac{1}{2}) = \frac{1}{8}$, and for $0 \leq y \leq t \leq 1$,

$P(Y \leq y|Y \leq t) = \frac{P(Y \leq y \cap Y \leq t)}{P(Y \leq t)} = \frac{P(Y \leq y)}{P(Y \leq t)} = \frac{y^3}{t^3} = (\frac{y}{t})^3$. □

Example 4-6:

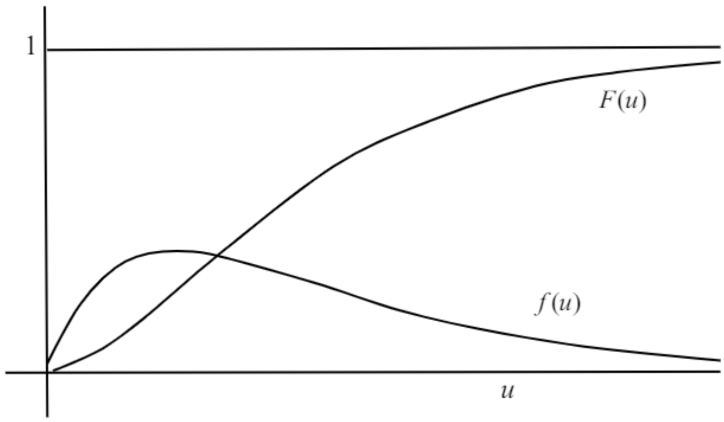
Continuous Random Variable on an Infinite Interval

U is a continuous random variable on the interval $(0,\infty)$ with density function.

$$f_U(u) = \begin{cases} ue^{-u} \text{ for } u>0 \\ 0 \text{ for } u\leq 0 \end{cases}.$$

The cdf is $F_U(u) = \int_{-\infty}^{u} f(t)\,dt = \int_0^u te^{-t}\,dt = -te^{-t} - e^{-t}\Big|_{t=0}^{t=u} = 1-(1+u)e^{-u}$ for $u>0$.

Then $F_U(u) = \begin{cases} 0 \text{ for } u\leq 0 \\ 1-(1+u)e^{-u}, \text{ for } u>0 \end{cases}$. $F(u)$ increases, approaching a limit of 1 as $u \to \infty$.



□

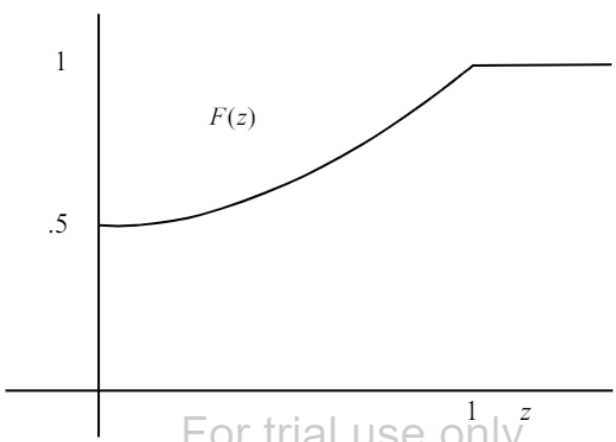
Example 4-7:

Mixed Random Variable

Z has a mixed distribution on the interval $[0,1)$. Z has probability of .5 at $Z=0$, and Z has density function $f_Z(z) = z$ for $0<z<1$, and Z has no density or probability elsewhere.

The cdf of Z is $F_Z(z) = \begin{cases} 0 \text{ if } z<0 \\ 0.5 \text{ if } z=0 \\ 0.5+\frac{1}{2}z^2 \text{ if } 0<z<1 \\ 1 \text{ if } z\geq 1 \end{cases}$

Note that there is a jump of (probability) 0.5 at $z=0$, and then $F(z)$ rises continuously on $(0,1)$.

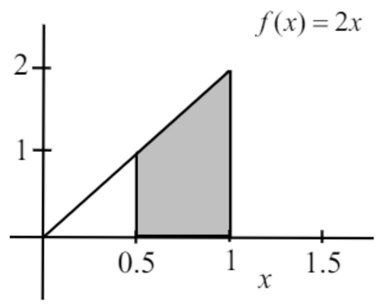


□

Some results and formulas relating to this section:

(i) For a continuous random variable X,
$P[a < X < b] = P[a \leq X < b] = P[a < X \leq b] = P[a \leq X \leq b] = \int_a^b f_X(x)\,dx$
so that when calculating the probability for a continuous random variable on an interval, it is irrelevant whether or not the endpoints are included. For the density function $f_X(x) = \begin{cases} 2x \text{ for } 0<x<1 \\ 0, \text{ otherwise} \end{cases}$, we have
$P[0.5 < X \leq 1] = \int_{0.5}^1 2x\,dx = x^2\Big|_{x=0.5}^{x=1} = 1 - (0.5)^2 = 0.75.$

This is illustrated in the shaded area in the graph below.



Also, for a continuous random variable, $P[X = a] = 0$, the probability at a single point is 0. Non-zero probabilities only exist over an interval, not at a single point.

(ii) For a continuous random variable, the hazard rate or failure rate is
$h(x) = \frac{f(x)}{1-F(x)} = -\frac{d}{dx}\,ln[1 - F(x)]$

(iii) If X has a mixed distribution, then $P[X = t]$ will be non-zero for some value(s) of t, and $P[a < X < b]$ will not always be equal to $P[a \leq X \leq b]$ (they will not be equal if X has a non-zero probability mass at either a or b).

(iv) $f(x)$ may be defined **piecewise**, meaning that $f(x)$ is defined by a different algebraic formula on different intervals. Example 4-13 below illustrates this.

(v) **Independence of random variables:** A more technical definition of independence of random variables will be given in a later section of these notes. One of the important consequences of random variables X and Y being independent is that
$P[(a < X \leq b) \cap (c < Y \leq d)] = P[a < X \leq b] \times P[c < Y \leq d]$
In general, what we mean by saying that random variables X and Y are independent is that if A is any event involving only X (such as $a < X \leq b$), and B is any event involving only Y, then A and B are independent events.

(vi) Conditional distribution of X given event A: Suppose that $f_X(x)$ is the density function or probability function of X, and suppose that A is an event. The conditional pdf or pf of

$$\text{"}X \text{ given } A\text{" is } f_{X|A}(x|A) = \begin{cases} \frac{f(x)}{P(A)} & \text{if } x \text{ is an outcome in event } A \\ 0 & \text{if } x \text{ is not an outcome in event } A \end{cases}$$

For example, suppose that $f_X(x) = \begin{cases} 2x \text{ for } 0<x<1 \\ 0, \text{ otherwise} \end{cases}$
and suppose that A is the event that $X \leq \frac{1}{2}$.

Then $P(A) = P(X \leq \frac{1}{2}) = \frac{1}{4}$, and for $0 < x \leq \frac{1}{2}$, $f_{X|A}(x|X \leq \frac{1}{2}) = \frac{2x}{1/4} = 8x$, and for $x > \frac{1}{2}$, $f_{X|A}(x|X \leq \frac{1}{2}) = 0$ (if we are given that $X \leq \frac{1}{2}$, then it is not possible for $x > \frac{1}{2}$, so the conditional density is 0 if $x > \frac{1}{2}$).

The conditional density must satisfy the same requirements as any probability density, it must integrate to 1 over its probability space. This is true for the example just presented, since $\int_0^{1/2} f_{X|A}(x|X \leq \frac{1}{2})\, dx = \int_0^{1/2} 8x\, dx = 1$.

Example 4-8:
A die is loaded in such a way that the probability of the face with j dots turning up is proportional to j for $j = 1, 2, 3, 4, 5, 6$. What is the probability, in one roll of the die, that an even number of dots will turn up?
Solution:
Let X denote the random variable representing the number of dots that appears when the die is rolled once.

Then, $P[X = k] = R \times k$ for $k = 1, 2, 3, 4, 5, 6$, where R is the proportional constant. Since the sum of all of the probabilities of points that can occur must be 1, it follows that $R \times [1 + 2 + 3 + 4 + 5 + 6] = 1$, so that $R = \frac{1}{21}$.

Then, $P[\text{even number of dots turns up}] = P[2] + P[4] + P[6] = \frac{2+4+6}{21} = \frac{4}{7}$. □

Example 4-9:
An ordinary single die is tossed repeatedly and independently until the first even number turns up. The random variable X is defined to be the number of the toss on which the first even number turns up. Find the probability that X is an even number.
Solution:
X is a discrete random variable that can take on an integer value of 1 or more. The probability function for X is $p(x) = P[X = x] = (\frac{1}{2})^x$ (this is the probability of $x - 1$ successive odd tosses followed by an even toss, we are using the independence of successive tosses). Then,

$P[X \text{ is even}] = P[2] + P[4] + P[6] + \cdots = (\frac{1}{2})^2 + (\frac{1}{2})^4 + (\frac{1}{2})^6 + \cdots = \frac{(\frac{1}{2})^2}{1-(\frac{1}{2})^2} = \frac{1}{3}$. □

Example 4-10:

Suppose that the continuous random variable X has density function

$f(x) = 3 - 48x^2$ for $-0.25 \le x \le 0.25$ (and $f(x) = 0$ elsewhere). Find $P[\frac{1}{8} \le X \le \frac{5}{16}]$.

Solution:

$P[0.125 \le X \le 0.3125] = P[0.125 \le X \le 0.25]$, since there is no density for X at points greater than 0.25. The probability is $\int_{0.125}^{0.25} (3 - 48x^2)\, dx = \frac{5}{32}$. □

Example 4-11:

Suppose that the continuous random variable X has the cumulative distribution function $F(x) = \frac{1}{1+e^{-x}}$ for $-\infty < x < \infty$. Find X's density function.

Solution: The density function for a continuous random variable is the first derivative of the distribution function. The density function of X is $f(x) = F'(x) = \frac{e^{-x}}{(1+e^{-x})^2}$. □

Example 4-12:

X is a random variable for which $P[X \le x] = 1 - e^{-x}$ for $x \ge 1$, and $P[X \le x] = 0$ for $x < 1$. Which of the following statements is true?

A) $P[X = 2] = 1 - e^{-2}$ and $P[X = 1] = 1 - e^{-1}$

B) $P[X = 2] = 1 - e^{-2}$ and $P[X \le 1] = 1 - e^{-1}$

C) $P[X = 2] = 1 - e^{-2}$ and $P[X < 1] = 1 - e^{-1}$

D) $P[X < 2] = 1 - e^{-2}$ and $P[X < 1] = 1 - e^{-1}$

E) $P[X < 2] = 1 - e^{-2}$ and $P[X = 1] = 1 - e^{-1}$

Solution:

Since $P[X \le x] = 1 - e^{-x}$ for $x \ge 1$, it follows that $P[X \le 1] = 1 - e^{-1}$.
But $P[X \le x] = 0$ if $x < 1$, and thus $P[X < 1] = 0$, so that $P[X = 1] = 1 - e^{-1}$
(since $P[X \le 1] = P[X < 1] + P[X = 1]$). This eliminates answers C and D. Since
the distribution function for X is continuous (and differentiable) for $x > 1$, it follows that $P[X = x] = 0$ for $x > 1$. This eliminates answers A, B and C. This is an example of a random variable X with a mixed distribution, a point of probability at $X = 1$, with $P(X = 1) = 1 - e^{-1}$, and a continuous distribution for $X > 1$ with pdf $f(x) = e^{-x}$ for $x > 1$. Answer: E □

Example 4-13:

A continuous random variable X has the density function

$$f(x) = \begin{cases} 2x & 0 < x < \frac{1}{2} \\ \frac{4-2x}{3} & \frac{1}{2} \le x < 2 \\ 0, & \text{elsewhere} \end{cases} . \quad \text{Find } \; P[0.25 < X \le 1.25]$$

Solution:

$$P[0.25 < X \le 1.25] = \int_{0.25}^{1.25} f(x)\,dx = \int_{0.25}^{0.5} 2x\,dx + \int_{0.5}^{1.25} \frac{4-2x}{3}\,dx = \frac{3}{4}$$

Note that since X is a continuous random variable, the probability $P[0.25 \le X < 1.25]$ would be the same as $P[.25 < X \le 1.25]$. This is an example of a density function defined piecewise. The only consequence of this is that in finding a probability for an interval that contains the point $\frac{1}{2}$, we must set up two integrals, one integral ending at right hand limit $\frac{1}{2}$, and the other integral starting at left hand limit $\frac{1}{2}$.

Also, note that if the density function was defined to be $g(x) = \begin{cases} 2x & 0 < x < \frac{1}{2} \\ 0 & x = 1/2 \\ \frac{4-2x}{3} & \frac{1}{2} < x \le 2 \end{cases}$

(0 density at $x = 1/2$), then all probabilities are unchanged (since the two density functions f and g differ at only one point, probability calculations, which are based on integrals of the density function over an interval, are the same for both f and g). □

Example 4-14:

The density function for the continuous random variable U is $f_U(u) = \begin{cases} e^{-u} \text{ for } u>0 \\ 0, \text{ for } u \le 0 \end{cases}$.

Find the probability $P[U \le 2|U > 1]$.

Solution:

$$P[U \le 2|U > 1] = \frac{P[(U \le 2) \cap (U > 1)]}{P[U > 1]} = \frac{P[1 < U \le 2]}{P[U > 1]}$$
$$P[1 < U \le 2] = \int_1^2 e^{-u}\,du = e^{-1} - e^{-2} , \; P[U > 1] = \int_1^{\infty} e^{-u}\,du = e^{-1}$$
$$P[U \le 2|U > 1] = \frac{e^{-1} - e^{-2}}{e^{-1}} = 1 - e^{-1}$$
□

Example 4-15:

An ordinary single die is tossed repeatedly until the first even number turns up. The random variable X is defined to be the number of the toss on which the first even number turns up. We define the following two events: $A = X$ is even , $B = X$ is a multiple of 3. Determine whether or not events A and B are independent.

Solution:

This is the same distribution as in Example 4-9. X is a discrete random variable that can take on an integer value of 1 or more. The probability function for X is $p(x) = P[X = x] = (\frac{1}{2})^x$

Then, $P[A] = P[X = 2 \text{ or } 4 \text{ or } 6 \ldots] = (\frac{1}{2})^2 + (\frac{1}{2})^4 + (\frac{1}{2})^6 + \cdots = (\frac{1}{2})^2 \times [\frac{1}{1-(\frac{1}{2})^2}] = \frac{1}{3}$

and $\quad P[B] = P[X = 3 \text{ or } 6 \text{ or } 9 \ldots] = (\frac{1}{2})^3 + (\frac{1}{2})^6 + (\frac{1}{2})^9 + \cdots = (\frac{1}{2})^3 \times [\frac{1}{1-(\frac{1}{2})^3}] = \frac{1}{7}$

$A \cap B = X$ is a multiple of 6 (multiple of 2 and of 3).

Then $P[A \cap B] = P[X = 6 \text{ or } 12 \text{ or } 18 \ldots]$

$$= (\tfrac{1}{2})^6 + (\tfrac{1}{2})^{12} + (\tfrac{1}{2})^{18} + \cdots = (\tfrac{1}{2})^6 \times [\tfrac{1}{1-(\frac{1}{2})^6}] = \tfrac{1}{63}$$

We note that $P[A \cap B] = \frac{1}{63} \neq \frac{1}{3} \times \frac{1}{7} = P[A] \times P[B]$, so thats A and B are not independent. □

Example 4-16:

A random sample of 4 independent random variables X_1, X_2, X_3, X_4 is obtained. Each of the X_i's has a density function of the form $\quad f(x) = \begin{cases} 2x \text{ for } 0<x<1 \\ 0, \text{ elsewhere} \end{cases}$

We define the following two random variables:

$Y = max\{X_1, X_2, X_3, X_4\}$ and $Z = min\{X_1, X_2, X_3, X_4\}$. Find the density functions of Y and Z.

Solution:

For Y we first find the distribution function.

$$\begin{aligned} P[Y \le y] &= P[max\{X_1, X_2, X_3, X_4\} \le y] \\ &= P[(X_1 \le y) \cap (X_2 \le y) \cap (X_3 \le y) \cap (X_4 \le y)] \\ &= P[X_1 \le y] \times P[X_2 \le y] \times P[X_3 \le y] \times P[X_4 \le y] = y^2 \times y^2 \times y^2 \times y^2 = y^8,\ 0 < y < 1 \end{aligned}$$

(We use the cdf of X, $P[X \le y] = \int_0^y 2x\,dx = y^2$.)

Thus, $F_Y(y) = P[Y \le y] = y^8 \rightarrow f_Y(y) = F_Y'(y) = 8y^7$ for $0 < y < 1$.

For Z we find the survival function (complement of the distribution function).

$$\begin{aligned} P[Z > z] &= P[min\{X_1, X_2, X_3, X_4\} > z] \\ &= P[(X_1 > z) \cap (X_2 > z) \cap (X_3 > z) \cap (X_4 > z)] \\ &= P[X_1 > z] \times P[X_2 > z] \times P[X_3 > z] \times P[X_4 > z] = (1 - z^2)^4 ,\ 0 < z < 1 \end{aligned}$$

Then $F_Z(z) = P[Z \le z] = 1 - P[Z > z] = 1 - (1 - z^2)^4$, $0 < z < 1$, and

$f_Z(z) = F_Z'(z) = 4(1 - z^2)^3(2z) = 8z(1 - z^2)^3$, $0 < z < 1$.

Y and Z are examples of order statistics on a collection of independent random variables. A little later we will consider order statistics in more detail. □

Example 4-17:

Example 4-4 considers the random variable
$X =$ number of successive independent tosses of a fair coin until the first head turns up.
X can be any integer ≥ 1, and the probability function of X is $p_X(x) = P[X = x] = \frac{1}{2^x}$
for $x = 1, 2, 3, \ldots$

(a) Find the probability function of the conditional distribution of X given that the first head occurs on an odd numbered toss. Find the probability that the first head occurs within the first 3 tosses given that the first head occurs on an odd numbered toss.

(b) Find the probability function of the conditional distribution of X given that the first head occurs within the first 5 tosses. Find the probability that the first head occurs within the first 3 tosses given that the first head occurs within the first 5 tosses.

Solution:

(a) A is the event that the first head occurs on an odd numbered toss.

$$P(A) = P[X = 1] + P[X = 3] + \cdots = \frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^5} + \frac{1}{2^7}\cdots$$
$$\frac{1}{2} \cdot [1 + \frac{1}{2^2} + (\frac{1}{2^2})^2 + (\frac{1}{2^2})^3 + \cdots] = \frac{1}{2} \times \frac{1}{1-(\frac{1}{2})^2} = \frac{2}{3}\,.$$

Then $p_{X|A}(x|X \text{ is odd}) = \frac{p_X(x)}{P(A)} = \frac{(\frac{1}{2})^x}{\frac{2}{3}} = \frac{3}{2} \times \frac{1}{2^x}$ if x is odd,

and $p_{X|A}(x|X \text{ is odd}) = 0$ if x is even. Then

$$P[X \leq 3|X \text{ is odd}] = p_{X|A}(1|X \text{ is odd}) + p_{X|A}(2|X \text{ is odd}) + p_{X|A}(3|X \text{ is odd})$$
$$\frac{3}{2} \times \frac{1}{2} + 0 + \frac{3}{2} \times \frac{1}{2^3} = 0.9375\,.$$

Note that we can also find $P[X \leq 3|X \text{ is odd}]$ using the definition of conditional probability;

$$P[X \leq 3|X \text{ is odd}] = \frac{P[X \leq 3 \cap X \text{ is odd}]}{P[X \text{ is odd}]} = \frac{P[X=1]+P[X=3]}{2/3} = \frac{3}{2} \times \frac{1}{2} + \frac{3}{2} \times \frac{1}{2^3}\,.$$

(b) B is the event that the first head occurs within the first 5 tosses.

$$P(B) = P[X \leq 5] = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \frac{1}{2^5} = \frac{31}{32}$$

$p_{X|B}(x|X \leq 5) = \frac{(\frac{1}{2})^x}{\frac{31}{32}}$ if $x = 1, 2, 3, 4, 5$, and $p_{X|B}(x|X \leq 5) = 0$ if $x > 5$

$$P[X \leq 3|X \leq 5] = p_{X|B}(1|X \leq 5) + p_{X|B}(2|X \leq 5) + p_{X|B}(3|X \leq 5)$$
$$= \frac{32}{31} \times \frac{1}{2} + \frac{32}{31} \times \frac{1}{2^2} + \frac{32}{31} \times \frac{1}{2^3} = 0\,.903$$

Alternatively,

$$P[X \leq 3|X \leq 5] = \frac{P[X \leq 3 \cap X \leq 5]}{P[X \leq 5]} = \frac{P[X=1]+P[X=2]+P[X=3]}{P[X \leq 5]} = \frac{\frac{1}{2}+\frac{1}{2^2}+\frac{1}{2^3}}{\frac{31}{32}} = 0.903$$ □

Example 4-18:

Bob just read a news report that suggested that one-quarter of all cars on the road are imports, and the rest are domestic. Bob decides to test this suggestion by watching the cars go by his house. Bob assumes that each successive car that goes by has a $\frac{1}{4}$ chance of being an import and a $\frac{3}{4}$ chance of being domestic. Bob knows cars, and he can tell the difference between imports and domestic cars. If Bob's assumption is correct, find the probability that Bob will see at least 2 imports pass his house before the 3rd domestic car passes his house.

Solution:

As soon as the 4th car passes his house, Bob will know whether or not at least 2 imports passed before the third domestic. If 2, 3 or 4 of the first 4 cars are imports, then the 2nd import passed his house before the 3rd domestic. If 0 or 1 of the first 4 cars are imports then the 3rd domestic passed his house before the 2nd import.

The probability of 2 of the first 4 cars being imports is the probability of any one of the following 6 successions of 4 cars occurring:

$$IIDD\ ,\ IDID\ ,\ IDDI\ ,\ DIID\ ,\ DIDI\ ,\ DDII\ .$$

Each one of those has a chance of $(\frac{1}{4})^2(\frac{3}{4})^2$ occurring, for a total probability of $6 \times \frac{9}{256} = \frac{27}{128}$.

The probability of 3 of the first 4 cars being imports is the probability of any one of the following 4 successions of 4 cars occurring:

$$IIID\ ,\ IIDI\ ,\ IDII\ ,\ DIII\ .$$

Each one of those has a chance of $(\frac{1}{4})^3 \times \frac{3}{4}$ occurring, for a total probability of $4 \times \frac{3}{256} = \frac{3}{64}$.

The probability that all 4 of the first 4 cars being imports is $(\frac{1}{4})^4 = \frac{1}{256}$.

Therefore, the overall total probability of at least 2 imports passing Bob's house before the 3rd domestic car passes his house is $\frac{27}{128} + \frac{3}{64} + \frac{1}{256} = \frac{67}{256}$. $\square$

PROBLEM SET 4

Random Variables and Probability Distributions

1. Let X be a discrete random variable with probability function $P[X=x] = \frac{2}{3^x}$ for $x = 1,2,3,\ldots$ What is the probability that X is even?

A) $\frac{1}{4}$ B) $\frac{2}{7}$ C) $\frac{1}{3}$ D) $\frac{2}{3}$ E) $\frac{3}{4}$

2. (SOA) In modeling the number of claims filed by an individual under an automobile policy during a three-year period, an actuary makes the simplifying assumption that for all integers $n \geq 0$, $p_{n+1} = \frac{1}{5}p_n$, where p_n represents the probability that the policyholder files n claims during the period. Under this assumption, what is the probability that a policyholder files more than one claim during the period?

A) 0.04 B) 0.16 C) 0.20 D) 0.80 E) 0.96

3. Let X be a continuous random variable with density function $f(x) = \begin{cases} 6x(1-x) & \text{for } 0<x<1 \\ 0 & \text{otherwise} \end{cases}$. Calculate $P[\,|X - \frac{1}{2}| > \frac{1}{4}]$.

A) 0.0521 B) 0.1563 C) 0.3125 D) 0.5000 E) 0.8000

4. Let X be a random variable with distribution function

$$F(x) = \begin{cases} 0 & \text{for } x<0 \\ \frac{x}{8} & \text{for } 0 \leq x < 1 \\ \frac{1}{4} + \frac{x}{8} & \text{for } 1 \leq x < 2 \\ \frac{3}{4} + \frac{x}{12} & \text{for } 2 \leq x < 3 \\ 1 & \text{for } x \geq 3 \end{cases}.$$

Calculate $P[1 \leq X \leq 2]$.

A) $\frac{1}{8}$ B) $\frac{3}{8}$ C) $\frac{7}{16}$ D) $\frac{13}{24}$ E) $\frac{19}{24}$

5. (SOA) In a small metropolitan area, annual losses due to storm, fire, and theft are independently distributed random variables. The pdf's are:

	Storm	Fire	Theft
$f(x)$	e^{-x}	$\frac{2e^{-2x/3}}{3}$	$\frac{5e^{-5x/12}}{12}$

Determine the probability that the maximum of these losses exceeds 3.

A) 0.002 B) 0.050 C) 0.159 D) 0.287 E) 0.414

6. Let X_1, X_2 and X_3 be three independent, identically distributed random variables each with density function $f(x) = \begin{cases} 3x^2 & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$. Let $Y = max\{X_1, X_2, X_3\}$. Find $P[Y > \frac{1}{2}]$.

A) $\frac{1}{64}$ B) $\frac{37}{64}$ C) $\frac{343}{512}$ D) $\frac{7}{8}$ E) $\frac{511}{512}$

7. Let the distribution function of X for $x > 0$ be $F(x) = 1 - \sum_{k=0}^{3} \frac{x^k e^{-x}}{k!}$.

What is the density function of X for $x > 0$?

A) e^{-x} B) $\frac{x^2 e^{-x}}{2}$ C) $\frac{x^3 e^{-x}}{6}$ D) $\frac{x^3 e^{-x}}{6} - e^{-x}$ E) $\frac{x^3 e^{-x}}{6} + e^{-x}$

8. Let X have the density function $f(x) = \frac{3x^2}{\theta^3}$ for $0 < x < \theta$, and $f(x) = 0$, otherwise.
If $P[X > 1] = \frac{7}{8}$, find the value of θ.

A) $\frac{1}{2}$ B) $(\frac{7}{8})^{1/3}$ C) $(\frac{8}{7})^{1/3}$ D) $2^{1/3}$ E) 2

9. (SOA) A group insurance policy covers the medical claims of the employees of a small company. The value, V, of the claims made in one year is described by $V = 100,000Y$ where Y is a random variable with density function $f(y) = \begin{cases} k(1-y)^4 & \text{for } 0 < y < 1 \\ 0 & \text{otherwise,} \end{cases}$ where k is a constant. What is the conditional probability that V exceeds 40,000, given that V exceeds 10,000?

A) 0.08 B) 0.13 C) 0.17 D) 0.20 E) 0.51

10. (SOA) An insurance company insures a large number of homes. The insured value, X of a randomly selected home is assumed to follow a distribution with density function

$$f(x) = \begin{cases} 3x^{-4} & \text{for } x > 1 \\ 0 & \text{otherwise.} \end{cases}$$

Given that a randomly selected home is insured for at least 1.5, what is the probability that it is insured for less than 2?

A) 0.578 B) 0.684 C) 0.704 D) 0.829 E) 0.875

11. (SOA) Two life insurance policies, each with a death benefit of 10,000 and a one-time premium of 500, are sold to a couple, one for each person. The policies will expire at the end of the tenth year. The probability that only the wife will survive at least ten years is 0.025, the probability that only the husband will survive at least ten years is 0.01, and the probability that both of them will survive at least ten years is 0.96. What is the expected excess of premiums over claims, given that the husband survives at least ten years?

A) 350 B) 385 C) 397 D) 870 E) 897

12. X_1 and X_2 are two independent random variables, but they have the same density function $f(x) = \begin{cases} 2x \text{ for } 0<x<1 \\ 0, \text{ elsewhere} \end{cases}$. Find the probability that the maximum of X_1 and X_2 is at least .5.

A) 0.92 B) 0.94 C) 0.96 D) 0.98 E) 1.00

13. For two random variables, the "distance" between two distributions is defined to be the maximum, $\underset{all\,x}{max}$ $|F_1(x) - F_2(x)|$ over the range for which F_1 and F_2 are defined, where $F(x)$ is the cumulative distribution function. Find the distance between the following two distributions:

(i) uniform on the interval $[0, 1]$

(ii) pdf is $f(x) = \frac{1}{(x+1)^2}$ for $0 < x < \infty$

A) 0 B) $\frac{1}{4}$ C) $\frac{1}{2}$ D) $\frac{3}{4}$ E) 1

14. A family health insurance policy pays the total of the first three claims in a year. If there is one claim during the year, the amount claimed is uniformly distributed between 100 and 500. If there are two claims in the year, the total amount claimed is uniformly distributed between 200 and 1000, and if there are three claims in the year, the total amount claimed is uniformly distributed between 500 and 2000. The probabilities of 0, 1, 2 and 3 claims in the year are $0.5, 0.3, 0.1, 0.1$, respectively. Find the probability that the insurer pays at least 500 in total claims for the year.

A) 0.10 B) 0.12 C) 0.14 D) 0.16 E) 0.18

15. (SOA) The loss due to a fire in a commercial building is modeled by a random variable X with density function

$$f(x) = \begin{cases} 0.005(20-x) & \text{for } 0 < x < 20 \\ 0 & \text{otherwise.} \end{cases}$$

Given that a fire loss exceeds 8, what is the probability that it exceeds 16?

A) $\frac{1}{25}$ B) $\frac{1}{9}$ C) $\frac{1}{8}$ D) $\frac{1}{3}$ E) $\frac{3}{7}$

16. (SOA) The lifetime of a machine part has a continuous distribution on the interval $(0, 40)$ with probability density function f, where $f(x)$ is proportional to $(10+x)^{-2}$.

Calculate the probability that the lifetime of the machine part is less than 6.

A) 0.04 B) 0.15 C) 0.47 D) 0.53 E) 0.94

17. (SOA) X is a continuous random variable with density function $f(x) = ce^{-x}$, $x > 1$.

Find $P[X < 3|X > 2]$

A) $1 - e^{-1}$ B) e^{-1} C) $1 - e^{-2}$ D) $e^{-1} - e^{-2}$ E) $e^{-2} - e^{-3}$

18. (SOA) Automobile policies are separated into two groups: low-risk and high-risk. Actuary Rahul examines low-risk policies, continuing until a policy with a claim is found and then stopping. Actuary Toby follows that same procedure with high-risk policies. Each low-risk policy has a 10% probability of having a claim. Each high-risk policy has a 20% probability of having a claim. The claim statuses of policies are mutually independent. Calculate the probability that Actuary Rahul examines fewer policies than Actuary Toby.

A) 0.2857 B) 0.3214 C) 0.3333 D) 0.3571 E) 0.4000

PROBLEM SET 4 SOLUTIONS

1. $P[X \text{ is even}] = P[X=2] + P[X=4] + P[X=6] + \cdots$
$$= \frac{2}{3} \times [\frac{1}{3} + \frac{1}{3^3} + \frac{1}{3^5} + \cdots] = \frac{2}{3^2} \times \frac{1}{1-\frac{1}{3^2}} = \frac{1}{4}$$ Answer: A

2. A requirement for a valid distribution is $\sum_{k=0}^{\infty} p_k = 1$.
Since $p_n = \frac{1}{5}p_{n-1} = \frac{1}{5} \times \frac{1}{5}p_{n-2} = \frac{1}{5} \cdot \frac{1}{5} \times \cdots \times \frac{1}{5}p_0 = (\frac{1}{5})^n \times p_0$, it follows that
$1 = \sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} (\frac{1}{5})^k \times p_0 = p_0 \times \frac{1}{1-\frac{1}{5}}$ (infinite geometric series) so that $p_0 = \frac{4}{5}$
and $p_k = (\frac{1}{5})^k \times \frac{4}{5}$. Then,
$P[N > 1] = 1 - P[N = 0 \text{ or } 1] = 1 - p_0 - p_1 = 1 - \frac{4}{5} - \frac{1}{5} \times \frac{4}{5} = \frac{1}{25}$. Answer: A

3. $P[\left|X - \frac{1}{2}\right| \le \frac{1}{4}] = P[-\frac{1}{4} \le X - \frac{1}{2} \le \frac{1}{4}] = P[\frac{1}{4} \le X \le \frac{3}{4}] = \int_{1/4}^{3/4} 6x(1-x)\,dx$
$$= 0.6875 \rightarrow P[\left|X - \frac{1}{2}\right| > \frac{1}{4}] = 1 - P[\left|X - \frac{1}{2}\right| \le \frac{1}{4}] = 0.3125$$ Answer: C

4. $P[1 \le X \le 2] = P[X \le 2] - P[X < 1] = F(2) - \lim_{x \to 1^-} F(x) = \frac{11}{12} - \frac{1}{8} = \frac{19}{24}$
Answer: E

5. $P[max\{S, F, T\} > 3] = 1 - P[max\{S, F, T,\} \le 3]$.
$$\begin{aligned} P[max\{S, F, T,\} \le 3] &= P[(S \le 3) \cap (F \le 3) \cap (T \le 3)] \\ &= P[S \le 3] \times P[F \le 3] \times P[T \le 3] \\ &= (1 - e^{-3/1})(1 - e^{-3/1.5})(1 - e^{-3/2.4}) = 0.586\,. \end{aligned}$$
$P[max\{S, F, T\} > 3] = 1 - 0.586 = 0.414$. Answer: E

6. $P[Y > \frac{1}{2}] = 1 - P[Y \le \frac{1}{2}] = 1 - P[(X_1 \le \frac{1}{2}) \cap (X_2 \le \frac{1}{2}) \cap (X_3 \le \frac{1}{2})]$
$$= 1 - (P[X \le \frac{1}{2}])^3 = 1 - [\int_0^{1/2} 3x^2 dx]^3 = 1 - (\frac{1}{8})^3 = \frac{511}{512}.$$ Answer: E

7. $f(x) = F'(x) = -\sum_{k=0}^{3} \frac{kx^{k-1}e^{-x} - x^k e^{-x}}{k!} = e^{-x} \times \sum_{k=0}^{3} \left[\frac{x^k - kx^{k-1}}{k!}\right]$
$$= e^{-x} \times \left[1 + \frac{x-1}{1} + \frac{x^2-2x}{2} + \frac{x^3-3x^2}{6}\right] = \frac{e^{-x}x^3}{6}$$ Answer: C

8. Since $f(x) = 0$ if $x > \theta$, and since $P[X > 1] = \frac{7}{8}$, we must conclude that $\theta > 1$.
Then, $P[X > 1] = \int_1^{\theta} f(x)\,dx = \int_1^{\theta} \frac{3x^2}{\theta^3}\,dx = 1 - \frac{1}{\theta^3} = \frac{7}{8}$, or equivalently, $\theta = 2$.
Answer: E

9. In order for $f(y)$ to be a properly defined density function it must be true that

$$\int_0^1 f(y)\,dy = \int_0^1 k(1-y)^4\,dy = 1 \rightarrow k \times \frac{1}{5} = 1 \rightarrow k = 5\,.$$

We wish to find the conditional probability $P[100,000Y > 40,000|100,000Y > 10,000]$.
For events A and B, the definition of the conditional probability $P[A|B]$ is $P[A|B] = \frac{P[A\cap B]}{P[B]}$.
With $A = 100,000Y > 40,000$ and $B = 100,000Y > 10,000$, we have $A \cap B = A$,
and, $P[100,000Y > 40,000|100,000Y > 10,000] = \frac{P[100,000Y>40,000]}{P[100,000Y>10,000]}$.
From the density function for Y we have
$P[100,000Y > 40,000] = P[Y > 0.4] = \int_{0.4}^1 f(y)\,dy = \int_{0.4}^1 5(1-y)^4\,dy = (0.6)^5$, and
$P[100,000Y > 10,000] = P[Y > 0.1] = \int_{0.1}^1 f(y)\,dy = \int_{0.1}^1 5(1-y)^4\,dy = (0.9)^5$.
The conditional probability in question is $\frac{(0.6)^5}{(0.9)^5} = 0.132$. Answer: B

10. We are asked to find a conditional probability $P[X < 2|X \geq 1.5]$.
The definition of conditional probability is $P[A|B] = \frac{P[A\cap B]}{P[B]}$.
Then, $P[X < 2|X \geq 1.5] = \frac{P[1.5\leq X<2]}{P[X\geq 1.5]}$.
From the given density function of X we get
$P[X \geq 1.5] = \int_{1.5}^{\infty} 3x^{-4}\,dx = \frac{1}{(1.5)^3} = 0.29630$ and
$P[1.5 \leq X < 2] = \int_{1.5}^{2} 3x^{-4}\,dx = \frac{1}{(1.5)^3} - \frac{1}{(2)^3} = 0.17130$.
Then, $P[X < 2|X \geq 1.5] = \frac{P[1.5\leq X<2]}{P[X\geq 1.5]} = \frac{0.17130}{0.29630} = 0.578$. Answer: A

11. W is the event that the wife will survive at least 10 years, and H is the event that the husband will survive at least 10 years. We are given $P[W \cap H'] = 0.025$, $P[W' \cap H] = 0.01$, and $P[W \cap H] = 0.96$.
Given that the husband survives at least 10 years, the probability that the wife survives at least 10 years is
$P[W|H] = \frac{P[W\cap H]}{P[H]} = \frac{0.96}{P[H]}$, and the probability that the wife does not survive at least 10 years is
$P[\overline{W}|H] = \frac{P[\overline{W}\cap H]}{P[H]} = \frac{0.01}{P[H]}$.
We can find $P[H] = P[W \cap H] + P[\overline{W} \cap H] = 0.96 + 0.01 = 0.97$, or use the table

		W		W'
$P[H]$ $= 0.97$	$\Leftarrow$	$P[W \cap H] = 0.96$ given	$+$	$P[W' \cap H] = 0.01$ given

Given that the husband survives 10 years, the claim will either be 0 if the wife survives 10 years, and 10,000 if the wife does not survive 10 years.

The expected amount of claim given that the husband survives 10 years is

$$0 \times P[W|H] + (10,000) \times P[\overline{W}|H] = 10,000 \times \frac{1}{97} = 103.09$$

The total premium is 1,000 (for the two insurance policies), so that the excess premium over expected claim is $1000 - 103.09 = 897$. Answer: E

12. $P[max\{X_1, X_2\} \geq .5] = 1 - P[(X_1 < 0.5) \cap (X_2 < 0.5)] = 1 - P[X_1 < 0.5] \times P[X_2 < 0.5]$
$P[X_1 < .5] = \int_0^{0.5} f(x)\,dx = \int_0^{0.5} 2x\,dx = 0.25$, and $P[X_2 < 0.5] = 0.25$ also.
Then $P[max\{X_1, X_2\} \geq 0.5] = 1 - 0.25 \times 0.25 = 0.9375$. Answer: B

13. $F_1(x) = x$ for $0 \leq x \leq 1$, and $F_1(x) = 1$ for $x > 1$.
$F_2(x) = \int_0^x \frac{1}{(t+1)^2}\,dt = 1 - \frac{1}{x+1}$ for $x > 0$.
For $0 \leq x \leq 1$, $|F_1(x) - F_2(x)| = |x + \frac{1}{x+1} - 1|$, which is maximized at either
$x = 0\,, 1$ or a critical point; critical points occur where $1 - \frac{1}{(x+1)^2} = 0$, or $x = 0$.
$|F_1(0) - F_2(0)| = 0$, $|F_1(1) - F_2(1)| = \frac{1}{2}$
For $x > 1$, $|F_1(x) - F_2(x)| = |1 + \frac{1}{x+1} - 1| = \frac{1}{x+1}$, which decreases.
The distance between the two distributions is $\frac{1}{2}$. Answer: C

14. $T =$ total claims for the year, $N =$ number of claims for the year .
$$P[T \geq 500] = \sum_{k=0}^{3} P[(T \geq 500) \cap (N = k)]$$
$$= P[(T \geq 500) \cap (N = 2)] + P[(T \geq 500) \cap (N = 3)]$$
(this is true since if there are 0 or 1 claim, then total must be $\leq$ 500).
$P[(T \geq 500) \cap (N = 2)] = P[T \geq 500|N = 2] \times P[N = 2] = \frac{1000-500}{1000-200} \times 0.1 = 0.0625$, and
$P[(T \geq 500) \cap (N = 3)] = P[T \geq 500|N = 3] \times P[N = 3] = 1 \times 0.1 = 0.1$.
Then $P[T \geq 500] = 0.0625 + 0.1 = 0.1625$. Answer: D

15. We are asked to find the conditional probability
$P[X > 16|X > 8] = \frac{P[X>16]}{P[X>8]}$
$P[X > 8] = \int_8^{20} 0.005 \times (20 - x)\,dx = 0.36$
$P[X > 16] = \int_{16}^{20} 0.005 \times (20 - x)\,dx = 0.04$
$P[X > 16|X > 8] = \frac{0.04}{0.36} = \frac{1}{9}$ Answer: B

16. $f(x) = c(10 + x)^{-2}$ for $0 < x < 40$
The total probability must be 1, so that $\int_0^{40} c(10 + x)^{-2}\,dx = c \times [\frac{1}{10} - \frac{1}{50}] = 1$.
Therefore, $c = 12.5$ and $f(x) = 12.5 \times (10 + x)^{-2}$.
Then, $P[X < 6] = \int_0^6 f(x)\,dx = \int_0^6 12.5 \times (10 + x)^{-2}\,dx$
$= -12.5 \times (10 + x)^{-1}\Big|_{x=0}^{x=6} = 0.46875$. Answer: C

17. $P[X<3|X>2]=\frac{P[2<X<3]}{P[X>2]}$

$P[X>2]=\int_2^\infty ce^{-x}\,dx=ce^{-2},\ P[2<X<3]=\int_2^3 ce^{-x}\,dx=c(e^{-2}-e^{-3})$

$P[X<3|X>2]=\frac{c(e^{-2}-e^{-3})}{ce^{-2}}=1-e^{-1}$

Note that we can find c, from $1=\int_1^\infty f(x)\,dx=\int_1^\infty ce^{-x}\,dx=ce^{-1}\rightarrow c=e$

but this is not necessary for this exercise. Answer: A

18. R denotes the number of policies that Rahul examines and T denotes the number of policies that Toby examines. Then $P(R=n)=(0.9)^{n-1}\times 0.1$ ($n-1$ policies with no claims followed by one policy with a claim). In a similar way, $P(T=m)=(0.8)^{m-1}\times 0.2$. We wish to find $P(R<T)$. We can use the probability rule $P(A)=\Sigma P(A|B_i)\times P(B_i)$, where $\{B_i\}$ forms a partition of a probability space. This is the Total Law of Probability reviewed on Page 65 of Section 2 of this study manual. Applying this rule, we get

$P(T>R)=\sum_{n=1}^{\infty}P(T>R|R=n)\times P(R=n)=\sum_{n=1}^{\infty}P(T>n)\times P(R=n)$

$P(T>n)=\sum_{m=n+1}^{\infty}P(T=m)=\sum_{m=n+1}^{\infty}(0.8)^{m-1}\times 0.2$

$=0.2\times[(0.8)^n+(0.8)^{n+1}+\cdots]=\frac{0.2\times(0.8)^n}{1-0.8}=(0.8)^n$

Then, $P(T>R)=\sum_{n=1}^{\infty}(0.8)^n\times(0.9)^{n-1}\times 0.1$

$=0.08\times\sum_{n=1}^{\infty}(0.8\times 0.9)^{n-1}=\frac{0.08}{1-0.72}=0.2857.$ Answer: A