Stevens Institute of Technology

SYS-601 Practice Exam #1 Solutions

1.1 Data Types and Terminology

(a) Consider the `airquality` data set which gives daily air quality measurements in New York from May to September 1973. It includes 154 observations of 4 variables:

Ozone	Ozone concentration (parts per billion)
Solar.R	Solar radiation (Langleys: 1 Langley = 41840 joules per square meter)
Wind	Average wind speed (miles per hour)
Temp	Maximum daily temperature (degrees Fahrenheit)

Select the best corresponding data type for each variable.

(i) Ozone

(A) Nominal (B) Ordinal (C) Interval (D) X Ratio

(ii) Solar.R

(A) Nominal (B) Ordinal (C) Interval (D) X Ratio

(iii) Wind

(A) Nominal (B) Ordinal (C) Interval (D) X Ratio

(iv) Temp

(A) Nominal (B) Ordinal (C) X Interval (D) Ratio

(b) *Select* the relationship known to exist between events X and Y if:

(i) $P(X \cup Y) = P(X) + P(Y)$

(A) Collectively Exhaustive (C) Independent
(B) Equally Likely (D) X Mutually Exclusive

(ii) $P(X) + P(Y) = 1$

(A) X Collectively Exhaustive (C) Independent
(B) Equally Likely (D) Mutually Exclusive

(iii) $P(X \cap Y) = P(X) \cdot P(Y)$

(A) Collectively Exhaustive
(B) Equally Likely
(C) X Independent
(D) Mutually Exclusive

(iv) $P(X) = P(Y)$

(A) Collectively Exhaustive
(B) X Equally Likely
(C) Independent
(D) Mutually Exclusive

(v) $P(X \cap Y) = 0$

(A) Collectively Exhaustive
(B) Equally Likely
(C) Independent
(D) X Mutually Exclusive

(vi) $P(X|Y) = P(X)$

(A) Collectively Exhaustive
(B) Equally Likely
(C) X Independent
(D) Mutually Exclusive

1.2 Descriptive Statistics

Consider the `warpbreaks` data set included which gives the number of warp breaks per loom, where a loom corresponds to a fixed length of yarn. It includes 54 observations of 3 variables:

breaks	The number of breaks
wool	The type of wool (A or B)
tension	The level of tension (low L, medium M, high H)

(a) *Select* the best corresponding data type for each variable:

(i) breaks

(A) Nominal (B) Ordinal (C) Interval (D) X Ratio

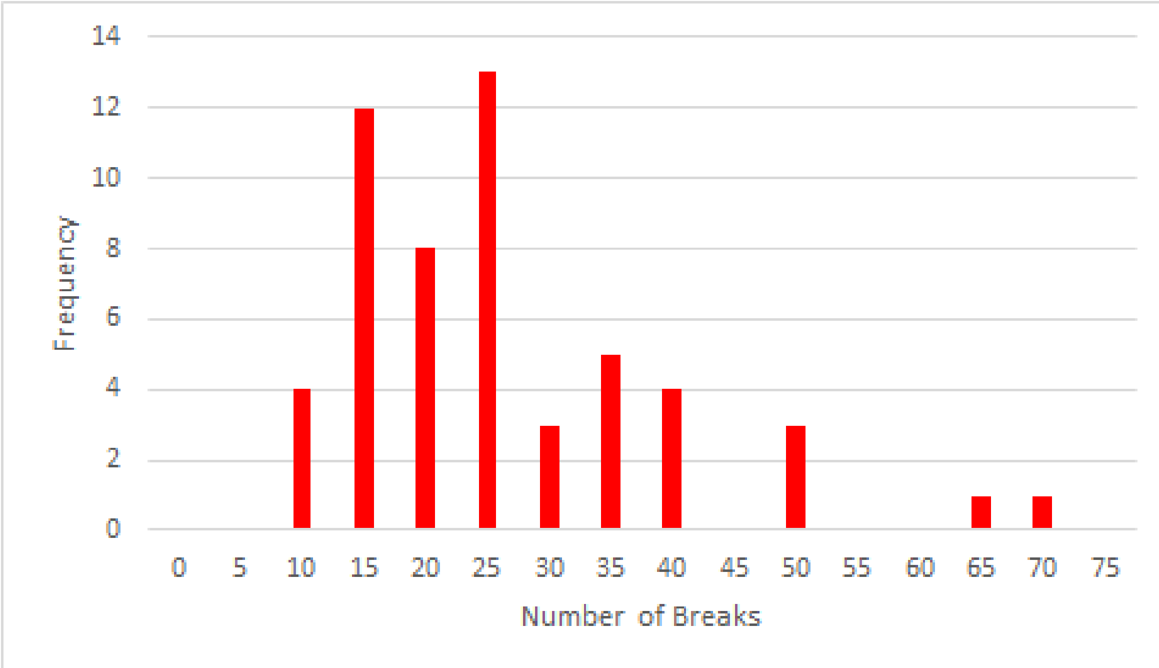
(ii) wool

(A) X Nominal (B) Ordinal (C) Interval (D) Ratio

(iii) tension

(A) Nominal (B) X Ordinal (C) Interval (D) Ratio

(b) Create a histogram showing the frequency of warp breaks in each loom (use the entire data set). Choose an appropriate bin size and label the x- and y-axes.



(c) What is the mean number of warp breaks? 28.15

(d) What is the median number of warp breaks? 26

(e) What is the sample standard deviation for number of warp breaks? 13.20

(f) What is the inter-quartile range for number of warp breaks?

Valid solutions include 15.75 (inclusive function) or 17.25 (exclusive function).

1.3 Probability Theory

Assume more than 20 warp breaks causes a quality control check to fail. The following frequency matrix shows the number of failed (F) and passed (P) tests for wool types A and B under low (L), medium (M), and high (H) tensions.

	A		B	
	F	P	F	P
L	9	0	6	3
M	5	4	7	2
H	6	3	3	6

(a) *Match* each probability expression with the best corresponding description:

$P(F)$	(i)	D	(A)	Conditional
$P(F \mid A)$	(ii)	A	(B)	Complement
$P(A \cap M)$	(iii)	C	(C)	Joint
$P(L \cup M)$	(iv)	E	(D)	Marginal
			(E)	Union

(b) What is the overall probability of failing a quality control test: $P(F)$? $\frac{36}{54} = 0.667$

(c) What is the probability of failing a quality control test given type A wool: $P(F|A)$? $\frac{20}{27} = 0.741$

(d) What is the probability of failing a quality control test given type B wool: $P(F|B)$? $\frac{16}{27} = 0.593$

(e) What is the probability of failing a quality control test given type A wool at medium tension: $P(F|A \cap M)$? $\frac{5}{9} = 0.556$

(f) What is the probability of failing a quality control test given type B wool at medium tension: $P(F|B \cap M)$? $\frac{7}{9} = 0.778$

1.4 Discrete Random Variables

The espresso machine at Café Java sometimes breaks down, requiring maintenance. There have been 3 such incidents over the past 2 years. Assume failures do not influence each other and the average failure rate is generally constant over time and consistent with prior observations.

The random variable X counts the number of espresso machine breakdowns in one year.

(a) *Select* the discrete distribution which best models the random variable X:

(A) Binomial

(B) Negative Binomial

(C) X Poisson

(D) Uniform

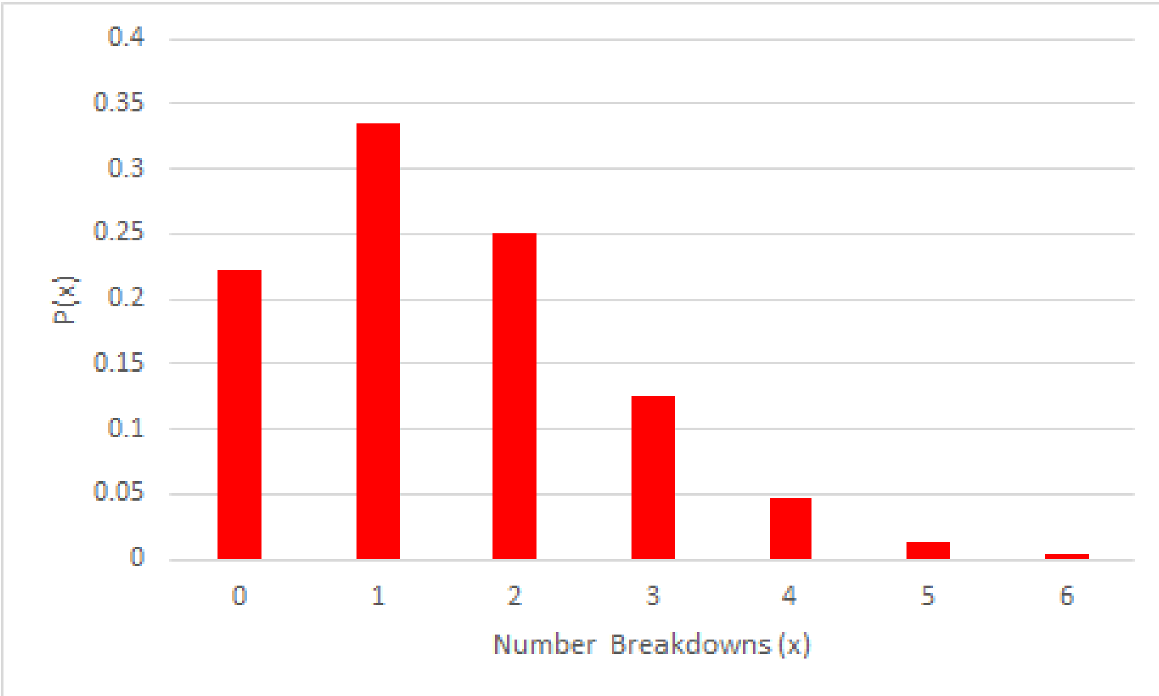
(b) Why might the assumptions required for this distribution not be suitable for real-life applications?

Key assumptions are independent failure events and constant failure rate. These may not be good assumptions if failures somehow influence each other or if the failure rate changes over time (for example, during higher burn-in and wear-out periods).

(c) What is the PMF equation $P\{X = x\} = P(x)$ for the distribution selected above? Substitute values for any required parameters.

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{\left(\frac{3}{2}\right)^x e^{-\frac{3}{2}}}{x!}$$

(d) Create a PMF plot for the random variable X for values between 0 and 6 breakdowns over a one year period. Label the x- and y-axes.



(e) What is the probability there will be more than one breakdown over the next year?

$$P\{X > 1\} = 1 - P(1) - P(0) = 1 - \frac{3}{2}e^{-\frac{3}{2}} - e^{-\frac{3}{2}} = 1 - 2.5e^{-1.5} = 0.442$$

(f) Assuming the espresso machine failed last week and was just repaired, what is the probability there will be no additional breakdowns over the next year?

It is irrelevant that the machine failed last week because we assume failures are independent and occur at a constant rate.

$$P(0) = e^{-\frac{3}{2}} = 0.223$$

(g) Which continuous distribution best models the amount of time between consecutive breakdowns?

(A) X Exponential

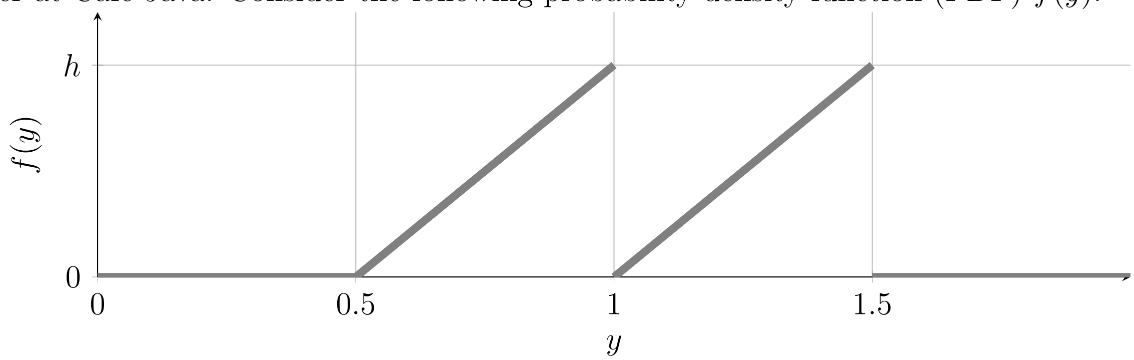
(B) Normal

(C) Uniform

(D) Weibull

1.5 Continuous Random Variables

The random variable Y measures the amount of time (in minutes) to prepare an espresso order at Café Java. Consider the following probability density function (PDF) $f(y)$:



(a) Using the definition of probability, what is the numerical value of h?

Using the definition that total probability must sum to 1, h can be solved by summing the area of the two triangles, each with base $(1.0 - 0.5)$ and height h.

$$1 = 2 \cdot \left(\frac{1}{2} \cdot (1.0 - 0.5) \cdot h \right) = 0.5 \cdot h \implies h = 2$$

(b) What is the PDF equation $f(y)$ for the region $0.5 \leq y < 1.0$?

This should be a linear equation which is 0 at $y = 0.5$ and h at $y = 1.0$.

$$f(y) = h \frac{(y - 0.5)}{(1.0 - 0.5)} = 2h(y - 0.5) = 4y - 2$$

(c) What is the expected value of Y?

The overall expected value is the weighted average of the center of area for each component triangle (center of area is at $\frac{a+2b}{3}$ for each).

$$E[Y] = \frac{1}{2} \frac{0.5 + 2 \cdot 1.0}{3} + \frac{1}{2} \frac{1.0 + 2 \cdot 1.5}{3} = \frac{2.5 + 4}{6} = \frac{13}{12} = 1.083$$

(d) What is the probability Y is greater than 1 minute: $P\{Y > 1.0\}$?

The point $y = 1.0$ splits the overall area in half; thus, $P\{Y > 1.0\} = 0.5$

(e) What is the probability Y is between 1 minute and 1.25 minutes: $P\{1.0 \leq Y < 1.25\}$?

The area of the triangle bounded by $1.0 \leq y \leq 1.25$ is:

$$P\{1.0 \leq Y < 1.25\} = \frac{1}{2} \cdot (1.25 - 1.0) \cdot \frac{h}{2} = \frac{1}{2} \cdot \frac{1}{4} \cdot 1 = \frac{1}{8} = 0.125$$

(f) What is the probability Y is between 1 and 1.25 minutes assuming Y is greater than 1 minute: $P\{1.0 \leq Y < 1.25 | Y > 1.0\}$?

Using the definition of conditional probability,

$$
\begin{aligned}
P\{1.0 \leq Y < 1.25 | Y > 1.0\} &= \frac{P\{1.0 \leq Y < 1.25 \cap Y > 1.0\}}{P\{Y > 1.0\}} \\
&= \frac{P\{1.0 < Y < 1.25\}}{P\{Y > 1.0\}} = \frac{0.125}{0.5} = 0.25
\end{aligned}
$$