Basic Probability 2

Stochastics

Illés Horváth

2021/09/14

Distributions

A random variable is a random number.

The distribution of a random variable is the possible values and their probabilities.

For a discrete variable X, the distribution can be described by the values

$$\mathbb{P}(X = k) = p_k, \quad k = 0, 1, \dots$$

Distributions

A random variable is a random number.
The distribution of a random variable is the possible values and their probabilities.
For a discrete variable X, the distribution can be described by the values

$$\mathbb{P}(X = k) = p_k, \quad k = 0, 1, \dots$$

For a continuous random variable X, its distribution is described by the cumulative distribution function (cdf)

$$F(x) = \mathbb{P}(X < x),$$

or by the probability density function

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}.$$

Bernoulli distribution

The same (or similar) distributions may arise from completely different random experiments. We are going to discuss several notable distributions.

Bernoulli distribution

The same (or similar) distributions may arise from completely different random experiments. We are going to discuss several notable distributions.

X has *Bernoulli distribution* with parameter p, or $X \sim \mathrm{I}(p)$ for short, if

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

X can be interpreted as the result of a single trial where the probability of success is p.

$$\mathbb{E}(X) = p.$$

Discrete uniform distribution

X has *discrete uniform distribution* with parameter n, or $X \sim \mathrm{DU}(n)$, if

$$\mathbb{P}(X = k) = \frac{1}{n}, \quad k = 1, \dots, n.$$

X can be interpreted as the result of a roll with a fair n-sided die.

$$\mathbb{E}(X) = \frac{1+n}{2}.$$

Geometric distribution

X has *geometric distribution* with parameter p, or $X \sim \mathrm{GEO}(p)$, if

$$\mathbb{P}(X = k) = p(1 - p)^{k-1}, \quad k = 1, 2, \ldots$$

X can be interpreted as the number of trials needed to get the first success, if each trial is independent and successful with probability p.

$$\mathbb{E}(X) = \frac{1}{p}.$$

Geometric distribution

X has *geometric distribution* with parameter p, or $X \sim \mathrm{GEO}(p)$, if

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots$$

X can be interpreted as the number of trials needed to get the first success, if each trial is independent and successful with probability p.

$$\mathbb{E}(X) = \frac{1}{p}.$$

Example. We keep rolling a fair 6-sided die until we roll a 6. The total number of rolls has distribution GEO(1/6).

Pessimistic geometric distribution

Y has *pessimistic geometric distribution* with parameter p, or $Y \sim \mathrm{PGEO}(p)$, if

$$\mathbb{P}(Y = k) = p(1-p)^k, \quad k = 0, 1, \ldots$$

Y can be interpreted as the number of trials *before* the first success, if each trial is independent and successful with probability p (so not counting the actual successful trial).

$$\mathbb{E}(Y) = \frac{1}{p} - 1.$$

Pessimistic geometric distribution

Y has *pessimistic geometric distribution* with parameter p, or $Y \sim \mathrm{PGEO}(p)$, if

$$\mathbb{P}(Y = k) = p(1-p)^k, \quad k = 0, 1, \ldots$$

Y can be interpreted as the number of trials *before* the first success, if each trial is independent and successful with probability p (so not counting the actual successful trial).

$$\mathbb{E}(Y) = \frac{1}{p} - 1.$$

If $X \sim \mathrm{GEO}(p)$, then $Y = X - 1 \sim \mathrm{PGEO}(p)$ and vice versa.

Binomial distribution

X has *binomial distribution* with parameters n and p, or $X \sim \mathrm{BIN}(n, p)$, if

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

$\left(\binom{n}{k} = \frac{n!}{k!(n-k)!}\right)$
X can be interpreted as the number of successful trials from n trials if each trial is independent and successful with probability p.

$$\mathbb{E}(X) = np.$$

Binomial distribution

X has *binomial distribution* with parameters n and p, or $X \sim \mathrm{BIN}(n, p)$, if

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

$\left(\binom{n}{k} = \frac{n!}{k!(n-k)!} \right)$
X can be interpreted as the number of successful trials from n trials if each trial is independent and successful with probability p.

$$\mathbb{E}(X) = np.$$

Example. If we flip a fair coin 10 times, the number of heads has distribution $\mathrm{BIN}(10, 1/2)$.

Poisson distribution

X has *Poisson distribution* with parameter λ, or $X \sim \mathrm{POI}(\lambda)$, if

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \ldots$$

X can be used to model the number of *rare events*, where the average number of events is λ. We assume that the events are coming from many different independent sources, and the contribution of each source is small.

$$\mathbb{E}(X) = \lambda.$$

Poisson distribution

Example. We know that on a low traffic road, on average 2 cars pass per minute. Then the number of cars passing in a given 1 minute interval has distribution $\mathrm{POI}(2)$.
Consider the following. The number of potential cars that can pass on the road in a given time interval is large, but the probability that a given car will pass there is very small; still, overall, the average number of cars is 2.

Poisson distribution

Example. We know that on a low traffic road, on average 2 cars pass per minute. Then the number of cars passing in a given 1 minute interval has distribution $\mathrm{POI}(2)$.
Consider the following. The number of potential cars that can pass on the road in a given time interval is large, but the probability that a given car will pass there is very small; still, overall, the average number of cars is 2.

Poisson distribution

Example. We know that on a low traffic road, on average 2 cars pass per minute. Then the number of cars passing in a given 1 minute interval has distribution $\mathrm{POI}(2)$.
Consider the following. The number of potential cars that can pass on the road in a given time interval is large, but the probability that a given car will pass there is very small; still, overall, the average number of cars is 2.

Example. The number of fires in a city in a given year has Poisson distribution.

Poisson distribution

Example. We know that on a low traffic road, on average 2 cars pass per minute. Then the number of cars passing in a given 1 minute interval has distribution $\mathrm{POI}(2)$.
Consider the following. The number of potential cars that can pass on the road in a given time interval is large, but the probability that a given car will pass there is very small; still, overall, the average number of cars is 2.

Example. The number of fires in a city in a given year has Poisson distribution.

Example. The number of packages arriving to an internet server in a given time interval has Poisson distribution.

Example. The number of errors in a book has Poisson distribution.

Uniform distribution

X has *uniform distribution* over the interval $[a, b]$, or $X \sim \mathrm{U}(a, b)$, if its pdf is

$$f(x) = \frac{1}{b-a}, \quad x \in [a, b].$$

This is a continuous distribution that can be used to model a random point within an interval.

$$\mathbb{E}(X) = \frac{a+b}{2}.$$

Exponential distribution

X has *exponential distribution* with parameter λ, or $X \sim \mathrm{EXP}(\lambda)$, if its pdf is

$$f(x) = \lambda e^{-\lambda x}, \quad x \in [0, \infty).$$

This is a continuous distribution that can be used to model the time of the first occurrence of a random event.

$$\mathbb{E}(X) = \frac{1}{\lambda}.$$

λ is the *rate* or density, so if λ is larger, X is typically smaller.

Exponential distribution

X has *exponential distribution* with parameter λ, or $X \sim \mathrm{EXP}(\lambda)$, if its pdf is

$$f(x) = \lambda e^{-\lambda x}, \quad x \in [0, \infty).$$

This is a continuous distribution that can be used to model the time of the first occurrence of a random event.

$$\mathbb{E}(X) = \frac{1}{\lambda}.$$

λ is the *rate* or density, so if λ is larger, X is typically smaller.

Examples. The time we have to wait for the first car to pass / first fire in a city / first request arriving to an internet server etc. has exponential distribution.

Normal distribution

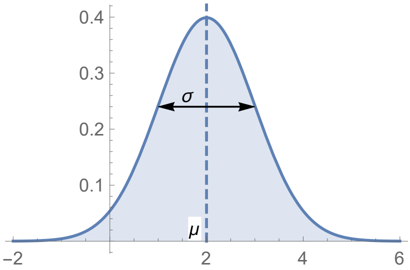
X has *normal distribution* with parameters μ and σ, or $X \sim \mathrm{N}(\mu, \sigma)$, if its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

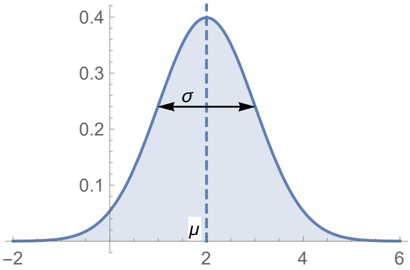
This is a continuous distribution that can be used to model a random number that is typically close to its mean, but can also be further away with smaller probability.

$$\mathbb{E}(X) = \mu, \qquad \mathbb{D}(X) = \sigma.$$

Normal distribution



Normal distribution



Example. The height (or other physical attributes) of a random person in a population can be modeled by normal distribution.
Example. Measurement error is often modeled by normal distribution.

Pareto distribution

X has *Pareto distribution* with parameters $A > 0$ (scale) and $\alpha > 0$ (shape), or $X \sim \mathrm{Pareto}(A, \alpha)$, if its cdf is

$$F(x) = 1 - \left(\frac{A}{x}\right)^{\alpha}, \quad x \geq A.$$

This is a continuous distribution that can be used to model random numbers where extremely large values may also occur.

$$\mathbb{E}(X) = \begin{cases} \frac{\alpha A}{\alpha - 1} & \alpha > 1 \\ \infty & 0 < \alpha \leq 1 \end{cases}$$

Pareto distribution

X has *Pareto distribution* with parameters $A > 0$ (scale) and $\alpha > 0$ (shape), or $X \sim \mathrm{Pareto}(A, \alpha)$, if its cdf is

$$F(x) = 1 - \left(\frac{A}{x}\right)^{\alpha}, \quad x \geq A.$$

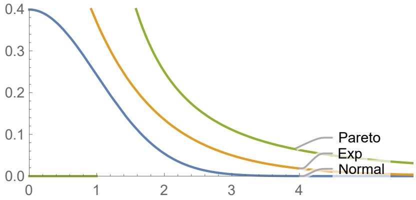
This is a continuous distribution that can be used to model random numbers where extremely large values may also occur.

$$\mathbb{E}(X) = \begin{cases} \frac{\alpha A}{\alpha - 1} & \alpha > 1 \\ \infty & 0 < \alpha \leq 1 \end{cases}$$

Example. The size of cities can be modeled by Pareto distribution.
Example. The distribution of wealth within a society can be modeled by Pareto distribution.

Decay of distributions

The pdf of the normal distribution decays very rapidly, the exponential distribution is still fast, Pareto is slower.



Two-dimensional distributions

Assume two discrete random variables X and Y are given on the same probability space. Then their *joint 2-dimensional distribution* can be described by

$$p_{k,l} = \mathbb{P}(X = k, Y = l), \quad k = 0, 1, \ldots, l = 0, 1, \ldots,$$

where the $p_{k,l}$'s are nonnegative and add up to 1.

Two-dimensional distributions

Assume two discrete random variables X and Y are given on the same probability space. Then their *joint 2-dimensional distribution* can be described by

$$p_{k,l} = \mathbb{P}(X = k, Y = l), \quad k = 0, 1, \ldots, l = 0, 1, \ldots,$$

where the $p_{k,l}$'s are nonnegative and add up to 1.

The *marginal distributions* of X and Y can be computed as

$$\mathbb{P}(X = k) = \sum_{l=0}^{\infty} \mathbb{P}(X = k, Y = l),$$

$$\mathbb{P}(Y = l) = \sum_{k=0}^{\infty} \mathbb{P}(X = k, Y = l).$$

Two-dimensional distributions

If X and Y are both continuous, then their *joint cumulative distribution function* is

$$F(x, y) = \mathbb{P}(X < x, Y < y),$$

and their *joint probability density function* is

$$f(x, y) = \frac{\partial^2 \mathbb{P}(X < x, Y < y)}{\partial x \partial y}.$$

Two-dimensional distributions

If X and Y are both continuous, then their *joint cumulative distribution function* is

$$F(x, y) = \mathbb{P}(X < x, Y < y),$$

and their *joint probability density function* is

$$f(x, y) = \frac{\partial^2 \mathbb{P}(X < x, Y < y)}{\partial x \partial y}.$$

The *marginal distributions* of X and Y have pdf's

$$f_X(x) = \int_{\mathbb{R}} f(x, y)\mathrm{d}y, \qquad f_Y(y) = \int_{\mathbb{R}} f(x, y)\mathrm{d}x.$$

Conditional distributions

If X and Y are discrete, then the *conditional distribution* of X assuming $Y = l$ is

$$\mathbb{P}(X = k | Y = l) = \frac{\mathbb{P}(X = k, Y = l)}{\mathbb{P}(Y = l)}.$$

Conditional distributions

If X and Y are discrete, then the *conditional distribution* of X assuming $Y = l$ is

$$\mathbb{P}(X = k | Y = l) = \frac{\mathbb{P}(X = k, Y = l)}{\mathbb{P}(Y = l)}.$$

If X and Y are continuous, then the conditional distribution of X assuming $Y = y$ has pdf

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

Independence

X and Y are *independent random variables* if the events $\{X \in A\}$ and $\{Y \in B\}$ are independent for any $A, B \subseteq \mathbb{R}$.

Independence

X and Y are *independent random variables* if the events $\{X \in A\}$ and $\{Y \in B\}$ are independent for any $A, B \subseteq \mathbb{R}$.

Theorem

X and Y are independent if and only if

$$\mathbb{P}(X = k, Y = l) = \mathbb{P}(X = k)\mathbb{P}(Y = l) \quad \forall k, l = 0, 1, \dots$$

for X, Y discrete and

$$f(x, y) = f_X(x) f_Y(y) \quad \forall x, y \in \mathbb{R}$$

for X, Y continuous.

Independence

X and Y are *independent random variables* if the events $\{X \in A\}$ and $\{Y \in B\}$ are independent for any $A, B \subseteq \mathbb{R}$.

Theorem

X and Y are independent if and only if

$$\mathbb{P}(X = k, Y = l) = \mathbb{P}(X = k)\mathbb{P}(Y = l) \quad \forall k, l = 0, 1, \ldots$$

for X, Y discrete and

$$f(x, y) = f_X(x) f_Y(y) \quad \forall x, y \in \mathbb{R}$$

for X, Y continuous.

If X and Y are independent, then $\mathbb{D}^2(X + Y) = \mathbb{D}^2(X) + \mathbb{D}^2(Y)$.

Expectation of functions

For a function $g(x, y)$,

$$\mathbb{E}(g(X,Y)) = \begin{cases} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} g(k,l)\mathbb{P}(X=k, Y=l) & \text{for } X, Y \text{ discrete,} \\ \int_{\mathbb{R}} \int_{\mathbb{R}} g(x,y) f(x,y) \mathrm{d}x \mathrm{d}y & \text{for } X, Y \text{ continuous.} \end{cases}$$

Covariance

The *covariance* of X and Y is

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Covariance measures linear dependence between X and Y.

Covariance

The *covariance* of X and Y is

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Covariance measures linear dependence between X and Y.

If $\mathrm{Cov}(X, Y) > 0$, then if X is large, then Y will also be typically large, and if X is small, then Y is typically also small.

Covariance

The *covariance* of X and Y is

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Covariance measures linear dependence between X and Y.

If $\mathrm{Cov}(X, Y) > 0$, then if X is large, then Y will also be typically large, and if X is small, then Y is typically also small.

If $\mathrm{Cov}(X, Y) < 0$, then if X is large, then Y will be typically small and vice versa.

Covariance

The *covariance* of X and Y is

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Covariance measures linear dependence between X and Y.

If $\mathrm{Cov}(X, Y) > 0$, then if X is large, then Y will also be typically large, and if X is small, then Y is typically also small.

If $\mathrm{Cov}(X, Y) < 0$, then if X is large, then Y will be typically small and vice versa.

If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$ (but the reverse is not true in general).

Correlation

The *correlation* of X and Y is

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\mathbb{D}(X)\mathbb{D}(Y)}.$$

Theorem (Cauchy-Schwarz inequality)

$$-1 \leq \mathrm{Corr}(X, Y) \leq 1.$$

The correlation is basically a normalized version of the covariance.

Correlation

The *correlation* of X and Y is

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\mathbb{D}(X)\mathbb{D}(Y)}.$$

Theorem (Cauchy-Schwarz inequality)

$$-1 \leq \mathrm{Corr}(X, Y) \leq 1.$$

The correlation is basically a normalized version of the covariance.

$\mathrm{Corr}(X, Y) = 1$ corresponds to full linear dependence between X and Y.

Problem 1

We throw a fair coin 5 times. What is the probability of getting two heads?

Problem 1

We throw a fair coin 5 times. What is the probability of getting two heads?

Solution. Let X denote the number of heads; the question is $\mathbb{P}(X = 2)$.

Problem 1

We throw a fair coin 5 times. What is the probability of getting two heads?

Solution. Let X denote the number of heads; the question is $\mathbb{P}(X = 2)$.

The distribution of X is $\mathrm{BIN}(5, 1/2)$, and so

$$\mathbb{P}(X = 2) = \binom{5}{2} \left(\frac{1}{2}\right)^2 \left(1 - \frac{1}{2}\right)^{5-2} = \frac{10}{32}.$$

Problem 5

There is an average of 2.3 shark attacks registered at the beaches of Florida each year. What is the probability that in a given year, at most 1 attack occurs?

Problem 5

There is an average of 2.3 shark attacks registered at the beaches of Florida each year. What is the probability that in a given year, at most 1 attack occurs?

Solution. Let X denote the number of shark attacks in the given year. Then the question is $\mathbb{P}(X \leq 1)$.

Problem 5

There is an average of 2.3 shark attacks registered at the beaches of Florida each year. What is the probability that in a given year, at most 1 attack occurs?

Solution. Let X denote the number of shark attacks in the given year. Then the question is $\mathbb{P}(X \leq 1)$.

The distribution of X is $X \sim \mathrm{POI}(2.3)$, so

$$\mathbb{P}(X \leq 1) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \underbrace{\frac{2.3^0}{0!}}_{=1} e^{-2.3} + \frac{2.3^1}{1!} e^{-2.3} \approx 0.331.$$

Problem 6

A book with 500 pages contains 1000 typos (errors). What is the probability that on a random page there are at least 2 typos? (We assume that each typo appears on every page with the same probability, and independently from other typos.)

Problem 6

A book with 500 pages contains 1000 typos (errors). What is the probability that on a random page there are at least 2 typos? (We assume that each typo appears on every page with the same probability, and independently from other typos.)

Solution. Let X denote the number of errors on the selected page. Since there are 1000 errors total in the book, and each error has a probability of $1/500$ to appear on that page,

Problem 6

A book with 500 pages contains 1000 typos (errors). What is the probability that on a random page there are at least 2 typos? (We assume that each typo appears on every page with the same probability, and independently from other typos.)

Solution. Let X denote the number of errors on the selected page. Since there are 1000 errors total in the book, and each error has a probability of $1/500$ to appear on that page, the distribution of X is $X \sim \mathrm{BIN}(1000, 1/500)$, and

$$\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) =$$

$$1 - \binom{1000}{0}\left(\frac{1}{500}\right)^0\left(\frac{499}{500}\right)^{1000} - \binom{1000}{1}\left(\frac{1}{500}\right)^1\left(\frac{499}{500}\right)^{999}.$$

Problem 6

On the other hand, 1000 errors on 500 pages means there are on average 2 errors per page.

Problem 6

On the other hand, 1000 errors on 500 pages means there are on average 2 errors per page. If there are on average 2 errors per page, then the number of errors on a random page has distribution $Y \sim \mathrm{POI}(2)$, and

$$\mathbb{P}(Y \geq 2) = 1 - \mathbb{P}(Y = 0) - \mathbb{P}(Y = 1) =$$
$$1 - \frac{2^0}{0!}e^{-2} - \frac{2^1}{1!}e^{-2}.$$

So which one is correct?

Problem 6

On the other hand, 1000 errors on 500 pages means there are on average 2 errors per page. If there are on average 2 errors per page, then the number of errors on a random page has distribution $Y \sim \mathrm{POI}(2)$, and

$$\mathbb{P}(Y \geq 2) = 1 - \mathbb{P}(Y = 0) - \mathbb{P}(Y = 1) =$$
$$1 - \frac{2^0}{0!}e^{-2} - \frac{2^1}{1!}e^{-2}.$$

So which one is correct?

Computing them numerically:

$$\mathbb{P}(X \geq 2) \approx 0.594265,$$
$$\mathbb{P}(Y \geq 2) \approx 0.593994.$$

Problem 6

In fact, this can be stated as a theorem.

Theorem

Let $n \to \infty$ and $p_n \to 0$ such that $np_n \to \lambda > 0$, and let $X_n \sim BIN(n, p_n)$ and $Y \sim POI(\lambda)$. Then

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(Y = k), \qquad \forall k \geq 0$$

(We also say that X_n converges in distribution to Y, or $X_n \xrightarrow{d} Y$.)

Problem 8

Assume that the age of a light bulb X (measured in 100 hours) has an exponential distribution such that $\mathbb{P}(X > 10) = 0.8$. Calculate the parameter of the exponential distribution and the mean of X.

Problem 8

Assume that the age of a light bulb X (measured in 100 hours) has an exponential distribution such that $\mathbb{P}(X > 10) = 0.8$. Calculate the parameter of the exponential distribution and the mean of X.

Solution. Let λ denote the parameter of the exponential distribution. Then its cdf is

$$F(x) = 1 - e^{-\lambda x},$$

and

$$\mathbb{P}(X > 10) = 1 - \mathbb{P}(X < 10) = 1 - F(10) = e^{-10\lambda} = 0.8,$$

from which $\lambda = -\log(0.8)/10 \approx 0.0223$, and

$$\mathbb{E}(X) = \frac{1}{\lambda} \approx 44.8 \quad \text{(in 100 hours)}.$$

Problem 10

In a class of 120 students, Stochastics and Calculus marks are as follows:

$C\backslash S$	1	2	3	4	5
1	1	2	2	1	4
2	2	4	4	8	2
3	4	8	8	12	8
4	5	4	6	9	6
5	0	6	4	6	4

We pick a student at random; let X denote his Stochastics mark and Y his Calculus mark.

a) $\mathbb{P}(\text{the student failed at least one of the courses}) =?$
b) $\mathbb{E}(X) =?$
c) $\mathbb{E}(X|Y \geq 4) =?$
d) Are X and Y independent?
e) $\mathrm{Cov}(X, Y) =?$

Problem 10

Solution.

a) A total of 22 students failed at least one of the courses (marked with red in the table), so

$$\mathbb{P}(\text{the student failed at least one of the courses}) = \frac{21}{120}.$$

$C \backslash S$	1	2	3	4	5
1	1	2	2	1	4
2	2	4	4	8	2
3	4	8	8	12	8
4	5	4	6	9	6
5	0	6	4	6	4

Problem 10

(b) We need to compute the marginal distribution of X.

$C \backslash S$	1	2	3	4	5
1	1	2	2	1	4
2	2	4	4	8	2
3	4	8	8	12	8
4	5	4	6	9	6
5	0	6	4	6	4
	12	24	24	36	24

So the marginal distribution of X is

k	1	2	3	4	5
$\mathbb{P}(X = k)$	$\frac{12}{120}$	$\frac{24}{120}$	$\frac{24}{120}$	$\frac{36}{120}$	$\frac{24}{120}$

and $\mathbb{E}(X) = \frac{12}{120} \cdot 1 + \frac{24}{120} \cdot 2 + \frac{36}{120} \cdot 3 + \frac{24}{120} \cdot 4 + \frac{24}{120} \cdot 5 = 3.2.$

Problem 10

(c) We need to compute the conditional distribution of X assuming $Y \geq 4$. First note that $\mathbb{P}(Y \geq 4) = \frac{50}{120}$.

$C \backslash S$	1	2	3	4	5
4	5	4	6	9	6
5	0	6	4	6	4
	5	10	10	15	10

Problem 10

(c) We need to compute the conditional distribution of X assuming $Y \geq 4$. First note that $\mathbb{P}(Y \geq 4) = \frac{50}{120}$.

$C \backslash S$	1	2	3	4	5
4	5	4	6	9	6
5	0	6	4	6	4
	5	10	10	15	10

So the conditional distribution of X assuming $Y \geq 4$ is

k	1	2	3	4	5
$\mathbb{P}(X = k \mid Y \geq 4)$	$\frac{5}{50}$	$\frac{10}{50}$	$\frac{10}{50}$	$\frac{15}{50}$	$\frac{10}{50}$

and $\mathbb{E}(X \mid Y \geq 4) = \frac{5}{50} \cdot 1 + \frac{10}{50} \cdot 2 + \frac{10}{50} \cdot 3 + \frac{15}{50} \cdot 4 + \frac{10}{50} \cdot 5 = 3.2.$

Problem 10

(d) No, for example

$$\mathbb{P}(X = 1, Y = 5) = 0 \neq \mathbb{P}(X = 1)\mathbb{P}(Y = 5) = \frac{12}{120} \cdot \frac{20}{120}.$$

Problem 10

(d) No, for example

$$\mathbb{P}(X=1, Y=5) = 0 \neq \mathbb{P}(X=1)\mathbb{P}(Y=5) = \frac{12}{120} \cdot \frac{20}{120}.$$

(e)

$$\mathbb{E}(X) = 3.2,$$
$$\mathbb{E}(Y) = 3.25,$$
$$\mathbb{E}(XY) = \frac{1}{120} \cdot 1 \cdot 1 + \frac{2}{120} \cdot 1 \cdot 2 + \cdots + \frac{4}{120} \cdot 5 \cdot 5 = 10.4,$$

so

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0,$$

even though X and Y are not independent.

Problem 10

(d) No, for example

$$\mathbb{P}(X = 1, Y = 5) = 0 \neq \mathbb{P}(X = 1)\mathbb{P}(Y = 5) = \frac{12}{120} \cdot \frac{20}{120}.$$

(e)

$$\mathbb{E}(X) = 3.2,$$
$$\mathbb{E}(Y) = 3.25,$$
$$\mathbb{E}(XY) = \frac{1}{120} \cdot 1 \cdot 1 + \frac{2}{120} \cdot 1 \cdot 2 + \dots + \frac{4}{120} \cdot 5 \cdot 5 = 10.4,$$

so

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0,$$

even though X and Y are not independent.

Bonus question: how was the table designed?