Exercise Solution

Below you'll see a table containing expectations and variances. Write the formula or shortcut for each one in the table. Where applicable, assume variables are independent.

Statistic	Shortcut or formula
$E(aX + b)$	$aE(X) + b$
$Var(aX + b)$	$a^2Var(X)$
$E(X)$	$\sum xP(X = x)$
$E(f(X))$	$\sum f(x)P(X = x)$
$Var(aX - bY)$	$a^2Var(X) + b^2Var(Y)$
$Var(X)$	$E(X - \mu)^2 = E(X^2) - \mu^2$
$E(aX - bY)$	$aE(X) - bE(Y)$
$E(X_1 + X_2 + X_3)$	$3E(X)$
$Var(X_1 + X_2 + X_3)$	$3Var(X)$
$E(X^2)$	$\sum x^2P(X = x)$
$Var(aX - b)$	$a^2Var(X)$

Exercise Solution

Sam likes to eat out at two restaurants. Restaurant A is generally more expensive than restaurant B, but the food quality is generally much better.

Below you'll find two probability distributions detailing how much Sam tends to spend at each restaurant. As a general rule, what would you say is the difference in price between the two restaurants? What's the variance of this?

Restaurant A:

x	**20**	**30**	**40**	**45**
P(X = x)	0.3	0.4	0.2	0.1

Restaurant B:

y	**10**	**15**	**18**
P(Y = y)	0.2	0.6	0.2

Let's start by finding the expectation and variance of X and Y.

$$E(X) = 20\times0.3 + 30\times0.4 + 40\times0.2 + 45\times0.1$$
$$= 6 + 12 + 8 + 4.5$$
$$= 30.5$$
$$Var(X) = (20-30.5)^2\times0.3 + (30-30.5)^2\times0.4 + (40-30.5)^2\times0.2 + (45-30.5)^2\times0.1$$
$$= (-10.5)^2\times0.3 + (-0.5)^2\times0.4 + 9.5^2\times0.2 + 14.5^2\times0.1$$
$$= 110.25\times0.3 + 0.25\times0.4 + 90.25\times0.2 + 210.25\times0.1$$
$$= 33.075 + 0.1 + 18.05 + 21.025$$
$$= 72.25$$

$$E(Y) = 10\times0.2 + 15\times0.6 + 18\times0.2$$
$$= 2 + 9 + 3.6$$
$$= 14.6$$
$$Var(Y) = (10-14.6)^2\times0.2 + (15-14.6)^2\times0.6 + (18-14.6)^2\times0.2$$
$$= (-4.6)^2\times0.2 + 0.4^2\times0.6 + 3.4^2\times0.2$$
$$= 21.16\times0.2 + 0.16\times0.6 + 11.56\times0.2$$
$$= 4.232 + 0.096 + 2.312$$
$$= 6.64$$

The difference between X and Y is modeled by X – Y.

$$E(X - Y) = E(X) - E(Y)$$
$$= 30.5 - 14.6$$
$$= 15.9$$

$$Var(X - Y) = Var(X) + Var(Y)$$
$$= 72.25 + 6.64$$
$$= 78.89$$

BULLET POINTS

- **Independent observations of X** are different instances of X. Each observation has the same probability distribution, but the outcomes can be different.
- If $X_1, X_2, ..., X_n$ are independent observations of X then:

  $$E(X_1 + X_2 + ... + X_n) = nE(X)$$
  $$Var(X_1 + X_2 + ... X_n) = nVar(X)$$

- If X and Y are independent random variables, then:

  $$E(X + Y) = E(X) + E(Y)$$
  $$E(X - Y) = E(X) - E(Y)$$
  $$Var(X + Y) = Var(X) + Var(Y)$$
  $$Var(X - Y) = Var(X) + Var(Y)$$

- The expectation and variance of linear transforms of X and Y are given by

  $$E(aX + bY) = aE(X) + bE(Y)$$
  $$E(aX - bY) = aE(X) - bE(Y)$$
  $$Var(aX + bY) = a2Var(X) + b2Var(Y)$$
  $$Var(aX - bY) = a2Var(X) + b2Var(Y)$$

BULLET POINTS

- The **geometric distribution** applies when you run a series of independent trials, there can be either a success or failure for each trial, the probability of success is the same for each trial, and the main thing you're interested in is how many trials are needed in order to get your first success.
- If the conditions are met for the geometric distribution, X is the number of trials needed to get the first successful outcome, and p is the probability of success in a trial, then

  $$X \sim Geo(p)$$

- The following probabilities apply if $X \sim Geo(p)$:

  $$P(X = r) = pq^{r-1}$$
  $$P(X > r) = q^r$$
  $$P(X \leq r) = 1 - q^r$$

- If $X \sim Geo(p)$ then

  $$E(X) = 1/p$$
  $$Var(X) = q/p^2$$

- The **binomial distribution** applies when you run a series of finite independent trials, there can be either a success or failure for each trial, the probability of success is the same for each trial, and the main thing you're interested in is the number of successes in the n independent trials.
- If the conditions are met for the binomial distribution, X is the number of successful outcomes out of n trials, and p is the probability of success in a trial, then

  $$X \sim B(n, p)$$

- If $X \sim B(n, p)$, you can calculate probabilities using

  $$P(X = r) = {}^nC_r\, p^r q^{n-r}$$

  where

  $${}^nC_r = \frac{n!}{r!\,(n - r)!}$$

- If $X \sim B(n, p)$, then

  $$E(X) = np$$
  $$Var(X) = npq$$

- The **Poisson distribution** applies when individual events occur at random and independently in a given interval, you know the mean number of occurrences in the interval or the rate of occurrences and this is finite, and you want to know the number of occurrences in a given interval.
- If the conditions are met for the Poisson distribution, X is the number of occurrences in a particular interval, and λ is the rate of occurrences, then

  $$X \sim Po(\lambda)$$

- If $X \sim Po(\lambda)$ then

  $$P(X = r) = \frac{e^{-\lambda} \lambda^r}{r!}$$
  $$E(X) = \lambda$$
  $$Var(X) = \lambda$$

- If $X \sim Po(\lambda_x)$, $Y \sim Po(\lambda_y)$ and X and Y are independent,

  $$X + Y \sim Po(\lambda_x + \lambda_y)$$

- If $X \sim B(n, p)$ where n is large and p is small, you can approximate it with $X \sim Po(np)$.

A quick guide to the geometric distribution

Here's a quick summary of everything you could possibly need to know about the Geometric distribution

When do I use it?

Use the Geometric distribution if you're running independent trials, each one can have a success or failure, and you're interested in how many trials are needed to get the first successful outcome

How do I calculate probabilities?

Use the following handy formulae. p is the probability of success in a trial, q = 1 - p, and X is the number of trials needed in order to get the first successful outcome. We say X ~ Geo(p).

$$P(X = r) = p\,q^{r-1}$$

The probability of the first success being in the r'th trial

$$P(X > r) = q^r$$

The probability you'll need more than r trials to get your first success

$$P(X \leq r) = 1 - q^r$$

The probability you'll need r trials or less to get your first success

What about the expectation and variance?

Just use the following

$$E(X) = 1/p \qquad Var(X) = q/p^2$$

Your quick guide to the binomial distribution

Here's a quick summary of everything you could possibly need to know about the binomial distribution

When do I use it?

Use the binomial distribution if you're running a fixed number of independent trials, each one can have a success or failure, and you're interested in the number of successes or failures

How do I calculate probabilities?

Use

$$P(X = r) = {}^nC_r\, p^r q^{n-r} \qquad {}^nC_r = \frac{n!}{r!\,(n-r)!}$$

where p is the probability of success in a trial, q = 1 - p, n is the number of trials, and X is the number of successes in the n trials.

What about the expectation and variance?

$$E(X) = np \qquad Var(X) = npq$$

Exercise Solution

It's time to test your statistical knowledge. Complete the table below, saying what normal distribution suits each situation, and what conditions there are.

Situation	Distribution	Condition
$X + Y$ **$X \sim N(\mu_x, \sigma^2_x)$, $Y \sim (\mu_y, \sigma^2_y)$**	$X + Y \sim N(\mu_x + \mu_y, \sigma^2_x + \sigma^2_y)$	X, Y are independent
$X - Y$ **$X \sim N(\mu_x, \sigma^2_x)$, $Y \sim (\mu_y, \sigma^2_y)$**	$X - Y \sim N(\mu_x - \mu_y, \sigma^2_x + \sigma^2_y)$	X, Y are independent
$aX + b$ **$X \sim N(\mu, \sigma^2)$**	$aX + b \sim N(a\mu + b, a^2\sigma^2)$	a, b are constant values
$X_1 + X_2 + ... + X_n$ **$X \sim N(\mu, \sigma^2)$**	$X_1 + X_2 + ... + X_n \sim N(n\mu, n\sigma^2)$	$X_1, X_2, ..., X_n$ are independent observations of X
Normal approximation of X **$X \sim B(n, p)$**	$X \sim N(np, npq)$	$np > 5$, $npq > 5$ Continuity correction required
Normal approximation of X **$X \sim Po(\lambda)$**	$X \sim N(\lambda, \lambda)$	$\lambda > 15$ Continuity correction required