prepared by Professor Jenny Baglivo

4 MATH442601|2 Notebook 4 **3**

4.1 Expected Value of a Random Variable 3

4.1.1 Definitions 3

4.1.2 Expectations for the Standard Models 6

4.2 Expected Value of a Function of a Random Variable 6

4.3 Properties of Expectation 10

4.4 Variance and Standard Deviation 10

4.4.1 Definitions and Properties 10

4.4.2 Chebyshev Inequality 13

4.4.3 Variances for the Standard Models 14

4.5 Expected Value of a Function of a Random Pair 14

4.5.1 Definitions and Properties 14

4.5.2 Covariance, Correlation, Association 18

4.5.3 Correlations for the Standard Models 23

4.5.4 Conditional Expectation, Regression 24

4.5.5 Historical Note: *Regression To The Mean* 29

4.6 Expected Value of a Linear Function of a Random k-Tuple 32

4.6.1 Mean, Variance, Covariance Matrix 32

4.6.2 Covariance 34

4.7 Combining Mutually Independent Random Variables 36

4.7.1 Random Sample, Sample Sum, Sample Mean 36

4.7.2 Mutually Independent Normal Random Variables 37

4.8 Moment Generating Functions 38

4 MATH442601|2 Notebook 4

This notebook is concerned with mathematical expectation. The notes correspond to material in Chapter 4 of the Rice textbook.

4.1 Expected Value of a Random Variable

The concept of expected value generalizes the concept of weighted average.

4.1.1 Definitions

1. <u>*Discrete Case:*</u> If X is a discrete random variable with range $\mathcal{R}$ and PMF $p(x)$, then the *expected value* of X is defined as follows:

$$E(X) = \sum_{x \in \mathcal{R}} x \; p(x),$$

as long as $\sum_{x \in \mathcal{R}} |x| \; p(x)$ converges. If the sum does not converge absolutely, then the expected value is said to be *indeterminate*.

2. <u>*Continuous Case:*</u> If X is a continuous random variable with range $\mathcal{R}$ and PDF $f(x)$, then the *expected value* of X is defined as follows:

$$E(X) = \int_{\mathcal{R}} x \; f(x) \; dx,$$

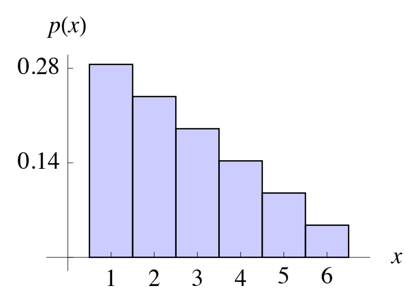
as long as $\int_{\mathcal{R}} |x| \; f(x) \; dx$ converges. If the integral does not converge absolutely, then the expected value is said to be *indeterminate*.

$E(X)$ is also called the *expectation* of X or the *mean* of X.

Exercise 1. Let X be the discrete random variable with PMF

$$p(x) = \frac{7-x}{21}, \quad \text{for } x = 1, 2, 3, 4, 5, 6;$$

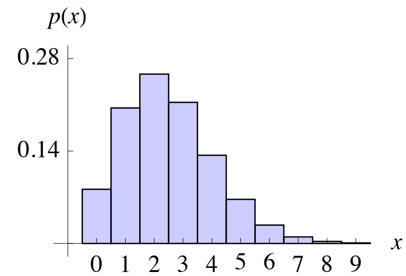
and 0 otherwise. Compute $E(X)$.



Exercise 2. Let X be the discrete random variable with PMF

$$p(x) = e^{-2.5}\frac{2.5^x}{x!}, \quad \text{for } x = 0, 1, 2, 3, \ldots;$$

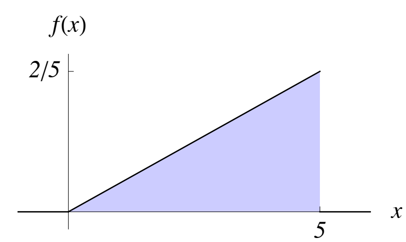
and 0 otherwise. Compute $E(X)$.



Exercise 3. Let X be the continuous random variable with PDF

$$f(x) = \frac{2x}{25}, \quad \text{for } x \in [0, 5];$$

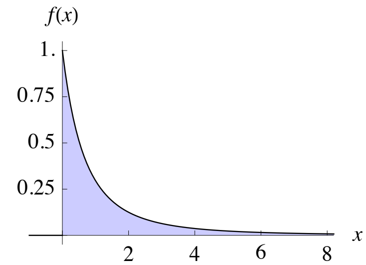
and 0 otherwise. Compute $E(X)$.



Exercise 4. (*a*) Let X be the continuous random variable with density function

$$f(x) = \frac{8}{(2+x)^3}, \quad \text{for } x \geq 0,$$

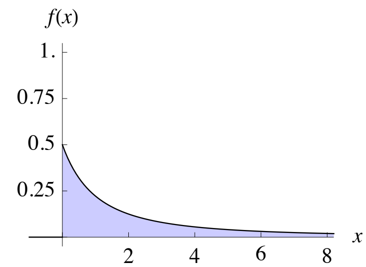
and 0 otherwise. Compute $E(X)$.



(*b*) Let X be the continuous random variable with PDF

$$f(x) = \frac{2}{(2+x)^2}, \quad \text{for } x \geq 0,$$

and 0 otherwise. Demonstrate that $E(X)$ is indeterminate.



4.1.2 Expectations for the Standard Models

The following table gives the expectations for the standard discrete and continuous models:

Probability Model:	***Expected Value:***
Discrete Uniform Distribution on $\{1, 2, \ldots, n\}$	$E(X) = \frac{n+1}{2}$
Hypergeometric Distribution based on choosing a subset of size n from an urn containing M special objects and a total of N objects	$E(X) = n\frac{M}{N}$
Bernoulli Distribution with success probability p	$E(X) = p$
Binomial Distribution based on n trials with success probability p	$E(X) = np$
Geometric Distribution where X is the trial number of the first success in a sequence of independent trials with success probability p	$E(X) = \frac{1}{p}$
Negative Binomial Distribution where X is the trial number of the r^{th} success in a sequence of independent trials with success probability p	$E(X) = \frac{r}{p}$
Poisson Distribution with parameter λ	$E(X) = \lambda$
(Continuous) Uniform Distribution on the interval (a, b)	$E(X) = \frac{a+b}{2}$
Exponential Distribution with parameter λ	$E(X) = \frac{1}{\lambda}$
Gamma Distribution with shape parameter α and scale parameter λ	$E(X) = \frac{\alpha}{\lambda}$
Cauchy Distribution with center a and spread b	$E(X)$ is indeterminate
Normal Distribution with mean μ and standard deviation σ	$E(X) = \mu$

4.2 Expected Value of a Function of a Random Variable

Let $g(X)$ be a real-valued function of the random variable X.

1. <u>*Discrete Case:*</u> If X is a discrete random variable with range $\mathcal{R}$ and PMF $p(x)$, then the *expected value* of $g(X)$ is defined as follows:

$$E(g(X)) = \sum_{x \in \mathcal{R}} g(x)\ p(x),$$

   as long as $\sum_{x \in \mathcal{R}} |g(x)|\ p(x)$ converges. If the sum does not converge absolutely, then the expected value is said to be *indeterminate*.

2. <u>*Continuous Case:*</u> If X is a continuous random variable with range $\mathcal{R}$ and PDF $f(x)$, then the *expected value* of $g(X)$ is defined as follows:

$$E(g(X)) = \int_{\mathcal{R}} g(x)\ f(x)\ dx,$$

   as long as $\int_{\mathcal{R}} |g(x)|\ f(x)\ dx$ converges. If the integral does not converge absolutely, then the expected value is said to be *indeterminate*.

Exercise. Assume that the probability of finding oil in a given drill hole is 0.30, and that the results (finding oil or not) are independent from drill hole to drill hole. An oil company drills one hole at a time. If they find oil, then they stop; otherwise, they continue. However, the company only has enough money to drill five holes. Let X be the number of holes drilled.

(a) Find the expected number of holes drilled, $E(X)$.

(b) The company has borrowed \$250,000 to buy equipment at the rate of 12% per drilling period, and have decided to pay back the loan once all drilling has been completed. Thus, $g(X) = 250(1.12)^X$ is the amount (in thousands of dollars) due once all drilling has been completed. Find the expected amount they will need to return, $E(g(X))$.

Exercise (Gambler's Ruin). Assume that the probability of winning a game is 0.50, and that the results (win or lose) are independent from game to game. A gambler decides to play the game until he or she wins. Let X be the number of games played.

(a) Find the expected number of games played, $E(X)$.

(b) The gambler places a \$1 bet on the first game, and, for each succeeding game, places a bet that is 3/2 the size of the previous bet:

$$1,\ \frac{3}{2},\ \left(\frac{3}{2}\right)^2,\ \ldots \quad \text{(until he or she wins).}$$

Thus, $g(X) = (3/2)^{X-1}$ is the amount (in dollars) bet on the last game played. Find the expected amount bet on the last game, $E(g(X))$.

(c) The gambler places a \$1 bet on the first game, and, for each succeeding game, doubles the bet placed of the previous game:

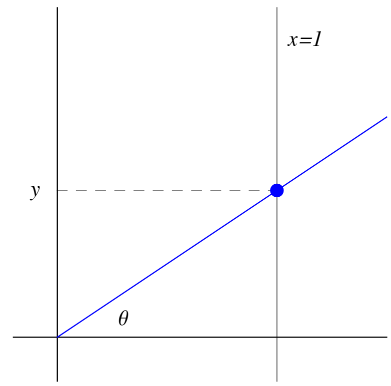
$$1,\ 2,\ 2^2,\ \ldots \quad \text{(until he or she wins).}$$

Thus, $g(X) = 2^{X-1}$ is the amount (in dollars) bet on the last game played. Demonstrate that the expected amount bet on the last game, $E(g(X))$, is indeterminate.

Exercise. Recall that positive angles are measured counterclockwise from the positive x-axis, and negative angles are measured clockwise.

Let Θ be a random angle in the interval $\left[0, \frac{\pi}{4}\right]$.

Given $\theta \in \left[0, \frac{\pi}{4}\right]$, construct a ray through the origin at angle θ and let $(1, y)$ be the point where the ray intersects the line $x = 1$, as illustrated in the plot to the right.



Assume Θ is a continuous uniform random variable on the interval $\left[0, \frac{\pi}{4}\right]$, and let $g(\Theta)$ be the y-coordinate of the point of intersection as described above.

(*a*) Find the expected y-coordinate, $E(g(\Theta))$.

(*b*) What value would you get if Θ were uniformly distributed on $\left[0, \frac{\pi}{2}\right)$?

4.3 Properties of Expectation

The following properties can be proven using properties of sums and integrals:

1. <u>*Constant Function:*</u> If a is a constant, then $E(a) = a$.
2. <u>*Linear Function:*</u> If $E(X)$ can be determined and a and b are constants, then
$$E(a + bX) = a + bE(X).$$
3. <u>*Linear Function of Functions:*</u> If $E(g_i(X))$ can be determined for $i = 1, 2, \ldots, k$, and a and b_i are constants for $i = 1, 2, \ldots, k$, then
$$E\left(a + \sum_{i=1}^{k} b_i g_i(X)\right) = a + \sum_{i=1}^{k} b_i E(g_i(X)).$$

For example, let X be an exponential random variable with parameter $\lambda = 1$. Integration by parts can be used to demonstrate that $E(X) = 1$, $E(X^2) = 2$ and $E(X^3) = 6$. Using these facts, we know that

$$E(5 + 3X - 4X^2 + X^3) = \underline{\hspace{6cm}}.$$

4.4 Variance and Standard Deviation

The mean of X is a measure of the *center* of the distribution. The variance and standard deviation are measures of the *spread* of the distribution.

4.4.1 Definitions and Properties

Let X be a random variable with mean $\mu = E(X)$. Then

1. The *variance* of X is defined as follows:
$$Var(X) = E((X - \mu)^2).$$
The notation $\sigma^2 = Var(X)$ is used to denote the variance.
2. The *standard deviation* of X is defined as follows:
$$SD(X) = \sqrt{Var(X)}.$$
The notation $\sigma = SD(X)$ is used to denote the standard deviation.

Exercise. Use the properties of expectation to demonstrate the following properties:

1. $Var(X) = E(X^2) - (E(X))^2$.
2. If $Y = a + bX$ for constants a and b, then $Var(Y) = b^2 Var(X)$ and $SD(Y) = |b| SD(X)$.

Exercise 1, page 3, continued. Find the mean and variance of the discrete random variable X given in Exercise 1, page 3.

Exercise 3, page 4, continued. Find the mean and variance of the continuous random variable X given in Exercise 3, page 4.

4.4.2 Chebyshev Inequality

Chebyshev's inequality gives us a lower bound for the probability that X is within k standard deviations of its mean, where k is a positive constant.

Theorem (Chebyshev Inequality). Let X be a random variable with mean $\mu = E(X)$ and standard deviation $\sigma = SD(X)$, and let k be a positive constant. Then

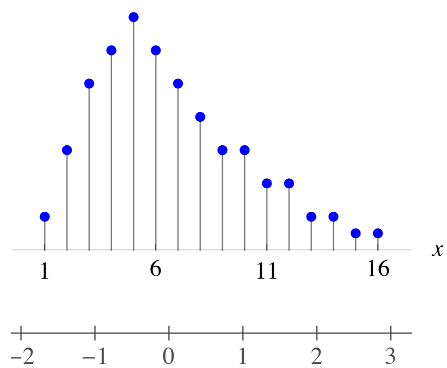
$$P(|X - \mu| < k\sigma) = P(\mu - k\sigma < X < \mu + k\sigma) > 1 - \frac{1}{k^2}.$$

Equivalently, we can say that $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$ when k is a positive constant.

For example, consider the discrete random variable X whose PMF is given on the left below.

x	$p(x)$
1	0.02
2	0.06
3	0.10
4	0.12
5	0.14
6	0.12
7	0.10
8	0.08

x	$p(x)$
9	0.06
10	0.06
11	0.04
12	0.04
13	0.02
14	0.02
15	0.01
16	0.01



For this random variable, $\mu = E(X) = 6.59$ and $\sigma = SD(X) = 3.33$.

To illustrate the Chebyshev inequality, let $k = 2$. Since

$\mu - 2\sigma =$ ______________________________, and $\mu + 2\sigma =$ ______________________________,

the probability that X is within 2 standard deviations of its mean is

$P(\mu - 2\sigma < X < \mu + 2\sigma) =$ ____.

This probability is greater than the guaranteed lower bound of _______________.

4.4.3 Variances for the Standard Models

The following table gives the variances for the standard discrete and continuous models:

Probability Model:	***Variance:***
Discrete Uniform Distribution on $\{1, 2, \ldots, n\}$	$Var(X) = \frac{n^2-1}{12}$
Hypergeometric Distribution with parameters n, M, N	$Var(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$
Bernoulli Distribution with success probability p	$Var(X) = p(1-p)$
Binomial Distribution based on n trials with success probability p	$Var(X) = np(1-p)$
Geometric Distribution with parameter p	$Var(X) = \frac{1-p}{p^2}$
Negative Binomial Distribution with parameters r, p	$Var(X) = \frac{r(1-p)}{p^2}$
Poisson Distribution with parameter λ	$Var(X) = \lambda$
(Continuous) Uniform Distribution on the interval (a, b)	$Var(X) = \frac{(b-a)^2}{12}$
Exponential Distribution with parameter λ	$Var(X) = \frac{1}{\lambda^2}$
Gamma Distribution with shape parameter α and scale parameter λ	$Var(X) = \frac{\alpha}{\lambda^2}$
Cauchy Distribution with center a and spread b	$Var(X)$ is indeterminate
Normal Distribution with mean μ and standard deviation σ	$Var(X) = \sigma^2$

4.5 Expected Value of a Function of a Random Pair

4.5.1 Definitions and Properties

Let $g(X, Y)$ be a real-valued function of the random pair (X, Y).

1. <u>*Discrete Case:*</u> If X and Y are discrete random variables with joint range $\mathcal{R} \subseteq \mathbf{R}^2$ and joint PMF $p(x, y)$, then the *expected value* of $g(X, Y)$ is defined as follows:

$$E(g(X,Y)) = \sum_{(x,y)\in\mathcal{R}} g(x,y)\; p(x,y),$$

   as long as $\sum_{(x,y)\in\mathcal{R}} |g(x,y)|\; p(x,y)$ converges. If the sum does not converge absolutely, then the expected value is said to be *indeterminate*.

2. <u>*Continuous Case:*</u> If X and Y are continuous random variables with joint range $\mathcal{R} \subseteq \mathbf{R}^2$ and joint PDF $f(x, y)$, then the *expected value* of $g(X, Y)$ is defined as follows:

$$E(g(X,Y)) = \iint_{\mathcal{R}} g(x,y)\; f(x,y)\; dA,$$

   as long as $\iint_{\mathcal{R}} |g(x,y)|\; f(x,y)\; dA$ converges. If the integral does not converge absolutely, then the expected value is said to be *indeterminate*.

Properties. The following properties can be proven using properties of sums and integrals:

1. *Linear Function of Functions:* If $E(g_i(X,Y))$ can be determined for $i = 1, 2, \ldots, k$, and a and b_i are constants for $i = 1, 2, \ldots, k$, then

$$E\left(a + \sum_{i=1}^{k} b_i g_i(X,Y)\right) = a + \sum_{i=1}^{k} b_i E(g_i(X,Y)).$$

2. *Product Function Under Independence:* If X and Y are independent random variables, and $g(X)$ and $h(Y)$ are real-valued functions, then

$$E(g(X)h(Y)) = E(g(X))E(h(Y)).$$

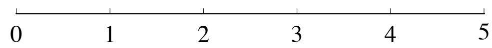
Exercise. Verify the 2^{nd} property when X is a continuous random variable with range $\mathcal{R}_X$ and PDF $f_X(x)$, and Y is a continuous random variable with range $\mathcal{R}_Y$ and PDF $f_Y(y)$.

Exercise. Let X and Y be the discrete random variables whose joint distribution is given in the table on the right.

Find $E(|X-Y|)$.

	$y=0$	$y=1$	$y=2$	$y=3$	$y=4$	*Sum:*
$x=0$	0.10	0.04	0.02	0.01	0.01	0.18
$x=1$	0.04	0.10	0.04	0.02	0.01	0.21
$x=2$	0.02	0.04	0.10	0.04	0.02	0.22
$x=3$	0.01	0.02	0.04	0.10	0.04	0.21
$x=4$	0.01	0.01	0.02	0.04	0.10	0.18
Sum:	0.18	0.21	0.22	0.21	0.18	1.00

Exercise. A stick of length 5 has a coordinate system as shown below:



The stick is broken at random in two places. Let X and Y be the locations of the two breaks and assume that X and Y are independent uniform random variables on the interval $(0, 5)$. Find the expected length of the middle segment.

4.5.2 Covariance, Correlation, Association

Let X and Y be random variables with finite means and finite variances. Then

1. The *covariance* of X and Y, denoted by $Cov(X, Y)$, is defined as follows:

$$Cov(X, Y) = E\left((X - \mu_x)(Y - \mu_y)\right),$$

   where $\mu_x = E(X)$ and $\mu_y = E(Y)$, provided the expectation exists.

2. The *correlation* of X and Y, denoted by $Corr(X, Y)$, is defined as follows:

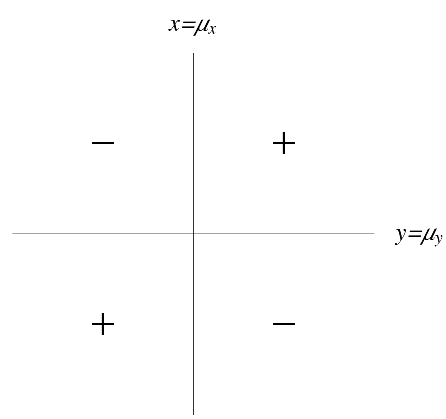
$$Corr(X, Y) = \frac{Cov(X, Y)}{SD(X)SD(Y)} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}},$$

   provided the covariance exists. Note that $\rho = Corr(X, Y)$ is a common shorthand for the correlation, and that ρ is often called the *correlation coefficient*.

<u>*To understand covariance,*</u>
imagine the plane with origin (μ_x, μ_y).

- Points in the 1^{st} and 3^{rd} quadrants contribute positive values, and
- points in the 2^{nd} and 4^{th} quadrants contribute negative values to the computation of covariance.

If points in the 1^{st} and 3^{rd} quadrants "outweigh" those in the second and fourth quadrants, then the covariance will be positive; the opposite will be true if the points in the 2^{nd} and 4^{th} quadrants "outweigh" those in the first and third quadrants.



<u>*Positive Association and Negative Association:*</u> Covariance and correlation are measures of the association between two random variables in the following sense:

1. X and Y are said to be *positively associated* if as X increases, Y tends to increase.

   If X and Y are positively associated, then $Cov(X, Y)$ and $Corr(X, Y)$ will be positive.

   For example, the height and weight of individuals in a population are positively associated; blood pressure and serum cholesterol levels are often positively associated.

2. X and Y are said to be *negatively associated* if as X increases, Y tends to decrease.

   If X and Y are negatively associated, then $Cov(X, Y)$ and $Corr(X, Y)$ will be negative.

   For example, educational level and indices of poor health are often negatively associated.

First list of properties. Properties of covariance and correlation include the following

1. $Cov(X, X) = Var(X)$.
2. $Cov(X, Y) = Cov(Y, X)$.
3. $Cov(X, Y) = E(XY) - E(X)E(Y)$.
4. $|Corr(X, Y)| \leq 1$.
5. $|Corr(X, Y)| = 1$ if and only if $Y = a + bX$, except possibly on a set of probability zero.
6. If X and Y are independent, then $Cov(X, Y) = 0$ and $Corr(X, Y) = 0$.

Exercise. Use properties of expectation to demonstrate that $Cov(X, Y) = E(XY) - E(X)E(Y)$.

Exercise. Let X and Y be the discrete random variables whose joint distribution is given in the table on the right.

Find $Cov(X,Y)$ and $Corr(X,Y)$.

	$y=0$	$y=1$	$y=2$	*Sum:*
$x=0$	0.05	0.05	0.10	0.20
$x=1$	0.15	0.10	0.07	0.32
$x=2$	0.30	0.10	0.08	0.48
Sum:	0.50	0.25	0.25	1.00

Exercise. Let X and Y be the continuous random variables whose joint PDF is

$$f(x,y) = \frac{2}{9}(y-x) \text{ when } 0 < x < y < 3; \text{ and } 0 \text{ otherwise.}$$

Find $Cov(X,Y)$ and $Corr(X,Y)$.

Correlated, uncorrelated. If $Corr(X, Y) = 0$, then X and Y are said to be *uncorrelated*; otherwise, they are said to be *correlated*. Independent random variables are uncorrelated, but uncorrelated random variables are *not* necessarily independent.

Exercise. Let X and Y be the discrete random variables whose joint distribution is given in the table on the right.

Demonstrate the X and Y are uncorrelated, but dependent.

	$y = 0$	$y = 1$	$y = 2$	*Sum:*
$x = 0$	0.05	0.10	0.05	0.20
$x = 1$	0.10	0.40	0.10	0.60
$x = 2$	0.05	0.10	0.05	0.20
Sum:	0.20	0.60	0.20	1.00

Second list of properties. Let a, b, c, and d be constants. Then

1. $Cov(a + bX, c + dY) = bdCov(X, Y)$.
2. $Corr(a + bX, c + dY) = \begin{cases} Corr(X, Y) & \text{when } bd > 0 \\ -Corr(X, Y) & \text{when } bd < 0 \end{cases}$

As a general example: Let X be the height of an individual in feet and let Y be the weight of the individual in pounds. If we change the measurement scales so that height was measured in inches (use $12X$ instead of X) and weight is measured in ounces (use $16Y$ instead of Y), then the covariance would change since

$$Cov(12X, 16Y) = (12)(16)Cov(X, Y),$$

but the correlation would remain the same since $Corr(12X, 16Y) = Corr(X, Y)$.

4.5.3 Correlations for the Standard Models

The following table gives the correlations for the standard bivariate models:

Probability Model:	***Correlation:***
Bivariate Hypergeometric Distribution with parameters n, (M_1, M_2, M_3)	$Corr(X, Y) = -\sqrt{\left(\frac{M_1}{M_2+M_3}\right)\left(\frac{M_2}{M_1+M_3}\right)}$
Trinomial Distribution with parameters n, (p_1, p_2, p_3)	$Corr(X, Y) = -\sqrt{\left(\frac{p_1}{1-p_1}\right)\left(\frac{p_2}{1-p_2}\right)}$
Bivariate Normal Distribution with parameters μ_x, μ_y, σ_x, σ_y, ρ	$Corr(X, Y) = \rho$

4.5.4 Conditional Expectation, Regression

Discrete case. Let X and Y be discrete random variables with joint PDF $p(x, y)$.

1. If $p_X(x) \neq 0$, then the *conditional expectation* (or the *conditional mean*) of Y given $X = x$, $E(Y|X = x)$, is defined as follows:

$$E(Y|X = x) = \sum_{y \in \mathcal{R}_{y|x}} y \; p_{Y|X=x}(y|x),$$

where $\mathcal{R}_{y|x}$ is the conditional range (the collection of y values with $p_{Y|X=x}(y|x) \neq 0$), provided the series converges absolutely.

2. If $p_Y(y) \neq 0$, then the *conditional expectation* (or the *conditional mean*) of X given $Y = y$ is defined as follows:

$$E(X|Y = y) = \sum_{x \in \mathcal{R}_{x|y}} x \; p_{X|Y=y}(x|y),$$

where $\mathcal{R}_{x|y}$ is the conditional range (the collection of x values with $(p_{X|Y=y}(x|y \neq 0)$, provided the series converges absolutely.

Continuous case. Let X and Y be continuous random variables with joint PDF $f(x, y)$.

1. If $f_X(x) \neq 0$, then the *conditional expectation* (or the *conditional mean*) of Y given $X = x$, $E(Y|X = x)$, is defined as follows:

$$E(Y|X = x) = \int_{\mathcal{R}_{y|x}} y \; f_{Y|X=x}(y|x) \; dy,$$

where $\mathcal{R}_{y|x}$ is the conditional range (the collection of y values with $f_{Y|X=x}(y|x) \neq 0$), provided the integral converges absolutely.

2. If $f_Y(y) \neq 0$, then the *conditional expectation* (or the *conditional mean*) of X given $Y = y$ is defined as follows:

$$E(X|Y = y) = \int_{\mathcal{R}_{x|y}} x \; f_{X|Y=y}(x|y) \; dx,$$

where $\mathcal{R}_{x|y}$ is the conditional range (the collection of x values with $(f_{X|Y=y}(x|y) \neq 0)$, provided the integral converges absolutely.

The formula for the conditional expectation of Y given $X = x$,

$$E(Y|X = x), \text{ as a function of } x,$$

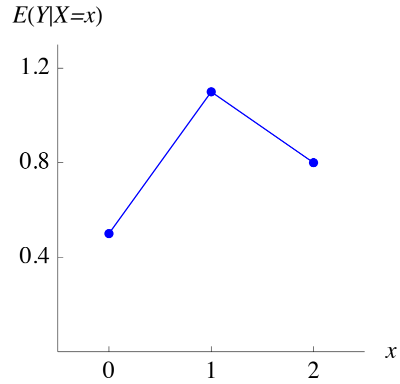
is often called the *regression* of Y on X.

Similarly, the formula for the conditional expectation $E(X|Y = y)$, as a function of y, is often called *regression* of X on Y.

Exercise. Let X and Y be the discrete random variables whose joint distribution is given in the following table:

	$y = 0$	$y = 1$	$y = 2$	*Sum:*
$x = 0$	0.210	0.030	0.060	0.30
$x = 1$	0.015	0.240	0.045	0.30
$x = 2$	0.200	0.080	0.120	0.40
Sum:	0.425	0.350	0.225	1.00

Find $E(Y|X = x)$, for $x = 0, 1, 2$.



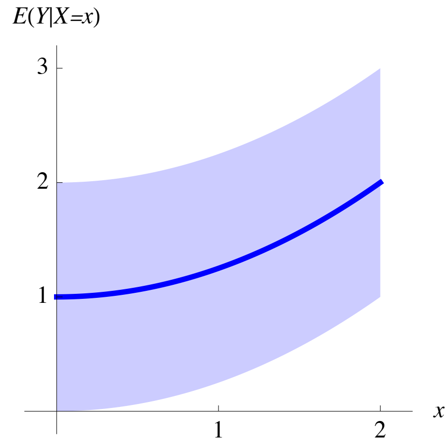
Exercise. Let X and Y be the continuous random variables whose joint PDF is

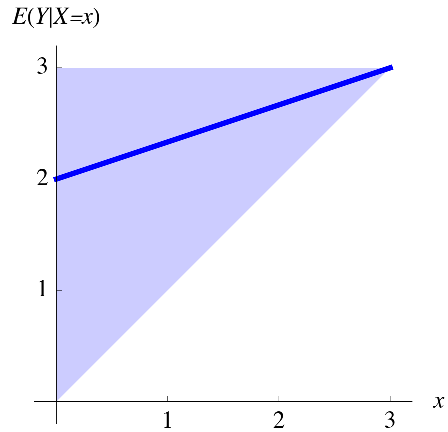
$$f(x,y) = \frac{1}{4} \text{ when } (x,y) \in \mathcal{R};\ 0 \text{ otherwise},$$

where $\mathcal{R} \subset \mathbf{R}^2$ is the region

$$\mathcal{R} = \{(x,y) \mid 0 \le x \le 2,\ x^2/4 \le y \le 2 + x^2/4\}.$$

Find $E(Y|X=x)$, for $0 \le x \le 2$.





Exercise. Let X and Y be the continuous random variables whose joint PDF is

$$f(x, y) = \frac{2}{9}(y - x) \text{ when } 0 < x < y < 3,$$

and 0 otherwise

Find $E(Y|X = x)$, for $0 < x < 3$.

When the conditional expectation of Y given $X = x$ is a linear function of x, then the formula has a particularly nice form:

Theorem (Linear Conditional Expectation). Let $\mu_x = E(X)$, $\mu_y = E(Y)$, $\sigma_x = SD(X)$, $\sigma_y = SD(Y)$ and $\rho = Corr(X, Y)$. If the conditional expectation $E(Y|X = x)$ is a linear function of x, then the form of this function must be

$$E(Y|X = x) = \mu_y + \frac{\rho\sigma_y}{\sigma_x}(x - \mu_x).$$

Exercise, continued. Let X and Y be the continuous random variables with joint PDF given on page 27. Use the work that we did on pages 21 and 27 to demonstrate that the conditional expectation of Y given $X = x$ satisfies the form given in the theorem above.

4.5.5 Historical Note: *Regression To The Mean*

The term "regression" was coined by the British scientist Sir Francis Galton in the late 1800's. Galton, and his young associate Karl Pearson, were interested in studying family resemblances. They used the bivariate normal distribution in their studies.[1]

Bivariate normal distribution. The bivariate normal distribution has five parameters,

$$\mu_x = E(X),\ \mu_y = E(Y),\ \sigma_x = SD(X),\ \sigma_y = SD(Y),\ \rho = Corr(X,Y),$$

and a complicated joint PDF $f(x,y)$.

A useful way to think about the joint PDF is as the product of a marginal PDF with a conditional PDF as follows:

$$f(x,y) = f_X(x)\ f_{Y|X=x}(y|x), \quad \text{for all real pairs } (x,y).$$

This is useful because

1. X has a normal distribution with mean μ_x and standard deviation σ_x, and
2. the conditional distribution of Y given $X = x$ is normal with parameters

$$E(Y|X=x) = \mu_y + \frac{\rho\sigma_y}{\sigma_x}(x-\mu_x) \ \text{ and } \ SD(Y|X=x) = \sigma_y\sqrt{1-\rho^2}.$$

Joint father-son height distribution. In one study, Galton and Pearson measured the heights of 1078 fathers and their grown sons and developed a bivariate normal distribution of father-son heights (in inches) with the following parameters:

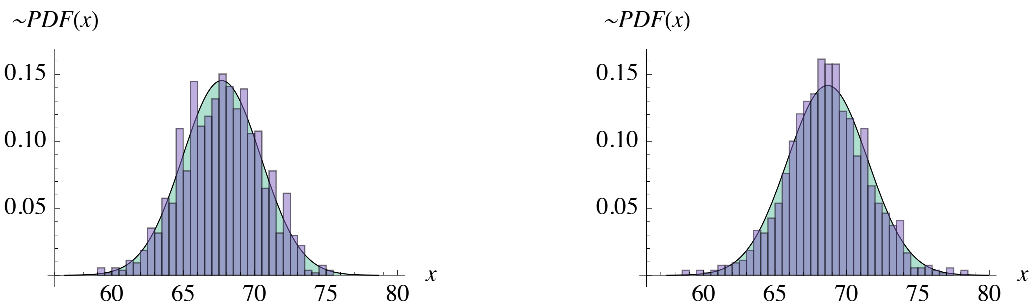
$$E(X) = 67.7,\ E(Y) = 68.7,\ SD(X) = 2.7,\ SD(Y) = 2.7,\ Corr(X,Y) = 0.50.$$

Further, the expected height in inches of a grown son whose father is x inches tall is

$$E(Y|X=x) = 68.7 + \frac{(0.50)(2.7)}{(2.7)}(x - 67.7) = 34.85 + 0.5x, \quad \text{for each } x.$$

[1] This historical note uses material from Chapter 25 of the textbook written by Freedman, Pisani and Purves, Norton & Company, 1991.

1. <u>*Marginal Models:*</u> Galton first examined the marginal distributions of fathers heights and sons heights:

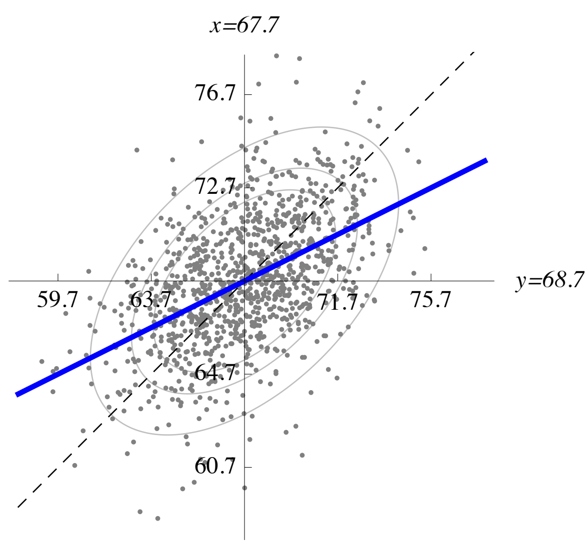


   (a) *Left Plot:* The left plot corresponds to the information for fathers heights. The normal model has mean 67.7 inches and standard deviation 2.7 inches.

   (b) *Right Plot:* The right plot corresponds to the information for sons heights. The normal model has mean 68.7 inches and standard deviation 2.7 inches.

   In each case, an *empirical histogram* of the heights is superimposed on a normal density curve; the shapes are similar.

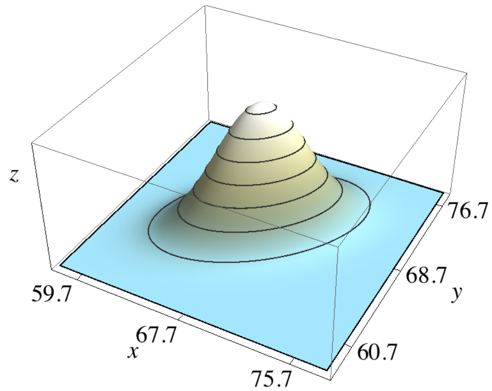
2. <u>*Enhanced scatter plot:*</u> Galton next examined a *scatter diagram* of the 1078 father-son height pairs:

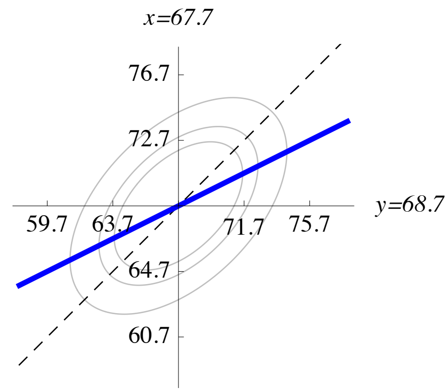


   Galton observed that the pattern of points was elliptical centered at the point of means and with major axis along the line $y = 1 + x$ (*dashed line*).

   Since sons are, on average, 1 inch taller than fathers and since the standard deviations of the distributions are equal, Galton initially thought that the line $y = 1 + x$ would represent the conditional expectation. Instead, he got $y = 34.85 + 0.5x$ (*solid line*).

3. *Galton-Pearson Model:* The observations above led Galton to propose using a bivariate normal model for the father-son height distribution (*left plot*):





For each x, the value of y on the line $y = 34.85 + 0.5x$ is midway between the values on

- the line $y = 1 + x$ and
- the horizontal line $y = 68.7$.

This lead Galton to observe that the conditional expected values had "regressed" toward the mean for all sons (*right plot*).

Probabilistic basis for regression to the mean. Lastly, after studying the works of Gregor Mendel[2] and Sir Francis Galton, the British statistician Sir Ronald Fisher[3] proved that Galton's idea of "regression to the mean" was a consequence of the probabilistic models of inheritance pioneered by Mendel.

[2]Gregor Mendel (1822-1884) is the Austrian-born Augustinian monk who is considered to be the father of modern genetics. He postulated the existence of entities now called *genes*, and developed simple chance models to explain how *genotypes* (the set of genes present in an individual) affect *phenotypes* (the observable physical characteristics of an individual).

[3]Ronald Fisher (1890-1962) is a founding father of the scientific discipline of statistics. His many achievements include the development of the principles of experimental design and the development of likelihood theory, which remain cornerstones of statistical practice to this day. Fisher developed likelihood theory as part of his work to put genetics on a firm mathematical foundation.

4.6 Expected Value of a Linear Function of a Random k-Tuple

In this section, we consider summarizing linear functions of the form

$$g(X_1, X_2, \ldots, X_k) = a + \sum_{i=1}^{k} b_i X_i,$$

where a and the b_i's are constants, and $(X_1, X_2, \ldots, X_k)$ is a random k-tuple.

4.6.1 Mean, Variance, Covariance Matrix

The mean and variance of a linear function of the X_i's can be written in terms of the means and variances of the X_i's, and the covariances $Cov(X_i, X_j)$ for $i \neq j$.

Theorem (Linear Function). Let $(X_1, X_2, \ldots, X_k)$ be a random k-tuple with finite means, variances and covariances, and let $V = a + \sum_{i=1}^{k} b_i X_i$ be a linear function of the X_i's, where a and the b_i's are constant. Then

1. $E(V) = a + \sum_{i=1}^{k} b_i E(X_i)$.
2. $Var(V) = \sum_{i=1}^{k} b_i^2 Var(X_i) + \sum_{i \neq j} b_i b_j Cov(X_i, X_j)$.

To demonstrate the second part, first note that

$$Var(V) = E\left((V - E(V))^2\right) = E\left(\left(\sum_{i=1}^{k} b_i (X_i - E(X_i))\right)^2\right).$$

Now (please complete),

Covariance matrix. Let $(X_1, X_2, \ldots, X_k)$ be a random k-tuple with finite means, variances and covariances. Let Σ be k-by-k matrix whose (i,j) term is $Cov(X_i, X_j)$. The matrix Σ is known as the *covariance matrix* of the random k-tuple.

For example, if $k = 4$, then we can write

$$\Sigma = [Cov(X_i, X_j)] = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & Cov(X_1, X_3) & Cov(X_1, X_4) \\ Cov(X_2, X_1) & Var(X_2) & Cov(X_2, X_3) & Cov(X_2, X_4) \\ Cov(X_3, X_1) & Cov(X_3, X_2) & Var(X_3) & Cov(X_3, X_4) \\ Cov(X_4, X_1) & Cov(X_4, X_2) & Cov(X_4, X_3) & Var(X_4) \end{bmatrix}$$

where I have substituted $Var(X_i) = Cov(X_i, X_i)$ along the main diagonal. Note that if the X_i's are mutually independent, then the covariance matrix is a diagonal matrix.

<u>*Computing Variances:*</u>

The covariance matrix gives us a convenient language for computing the variance of a linear function of the X_i's. Specifically, let $V = a + \sum_{i=1}^{k} b_i X_i$ be a linear function of the X_i's, and let $\boldsymbol{b}$ be the column vector of the coefficients of the X_i's in the formula for V. Then

$$Var(V) = \boldsymbol{b}^T \Sigma \boldsymbol{b} = \sum_{i=1}^{k} b_i^2 Var(X_i) + \sum_{i \neq j} b_i b_j Cov(X_i, X_j),$$

where $\boldsymbol{b}^T$ is the row vector whose terms are the coefficients of the X_i's in the formula for V. In applied problems, you may find it useful to obtain $Var(V)$ by computing $\boldsymbol{b}^T \Sigma \boldsymbol{b}$.

Exercise. Suppose that the random triple (X, Y, Z) has the following summaries:

$$E(X) = E(Y) = E(Z) = 10 \quad \text{and} \quad Var(X) = Var(Y) = Var(Z) = 4,$$

and let $V = 1 + 2X - 3Y + 4Z$.

(a) The expected value of V is ______________________________.

(b) If X, Y and Z are mutually independent, find the variance of V.

(c) If $Corr(X, Y) = Corr(X, Z) = Corr(Y, Z) = \frac{1}{2}$, find the variance of V.

4.6.2 Covariance

The covariance between two linear functions of the X_i's can be written in terms of the means and variances of the X_i's, and the covariances $Cov(X_i, X_j)$ for $i \neq j$.

Theorem (Covariance). Let $(X_1, X_2, \ldots, X_k)$ be a random k-tuple with finite means, variances and covariances, and let

$$V = a + \sum_{i=1}^{k} b_i X_i \quad \text{and} \quad W = c + \sum_{i=1}^{k} d_i X_i$$

be linear functions of the X_i's, where a, c, and the b_i's and d_i's are constants. Then

$$Cov(V, W) = \boldsymbol{b}^T \Sigma \boldsymbol{d} = \sum_{i=1}^{k} b_i d_i Var(X_i) + \sum_{i \neq j} b_i d_j Cov(X_i, X_j),$$

where $\boldsymbol{b}$ and $\boldsymbol{d}$ are the column vectors of coefficients of the X_i's for V and W, respectively, and Σ is the covariance matrix of the random k-tuple $(X_1, X_2, \ldots, X_k)$.

Exercise. Suppose that the random triple (X, Y, Z) has the following summaries:

$$Var(X) = Var(Y) = Var(Z) = 4 \text{ and } Corr(X, Y) = Corr(X, Z) = Corr(Y, Z) = \frac{1}{2},$$

and let $V = 1 + 2X - 3Y + 4Z$ and $W = 2 + X + Y - Z$.

Find $Cov(V, W)$ and $Corr(V, W)$.

4.7 Combining Mutually Independent Random Variables

Let X_1, X_2, ..., X_k be mutually independent random variables. Then

1. <u>*Linear Function:*</u> If $V = a + \sum_{i=1}^{k} b_i \; g_i(X_i)$, then

$$E(V) = a + \sum_{i=1}^{k} b_i E(g_i(X_i)) \quad \text{and} \quad Var(V) = \sum_{i=1}^{k} b_i^2 Var(g_i(X_i))$$

   when the means and variances on the right of each equality can be determined.

2. <u>*Product Function:*</u> If $W = \prod_{i=1}^{k} g_i(X_i)$, then

$$E(W) = \prod_{i=1}^{k} E(g_i(X_i)),$$

   when each expectation on the right can be determined.

4.7.1 Random Sample, Sample Sum, Sample Mean

A *random sample* of size k from the X distribution is a list X_1, X_2, ..., X_k, of k mutually independent random variables, each with the same distribution as X.

Given a random sample of size k,

1. The *sample sum* is the quantity $S = \sum_{i=1}^{k} X_i$.
2. The *sample mean* is the quantity $\overline{X} = \frac{1}{k} \sum_{i=1}^{k} X_i$.

Exercise. Let X_1, X_2, ..., X_k be a random sample from a distribution with mean $\mu = E(X)$ and standard deviation $\sigma = SD(X)$. Find expressions for the mean and variance of the sample sum, and for the mean and variance of the sample mean.

4.7.2 Mutually Independent Normal Random Variables

A linear function of independent normal random variables is itself a normal random variable.

> **Theorem (Normal Random Variables).** Let X_1, X_2, $\ldots$, X_k be mutually independent normal random variables, and let $V = a + \sum_{i=1}^{k} b_i X_i$, where a and the b_i's are constants. Then V is a normal random variable with
>
> $$E(V) = a + \sum_{i=1}^{k} b_i E(X_i) \text{ and } Var(V) = \sum_{i=1}^{k} b_i^2 Var(X_i).$$

Exercise. The Big Target Corporation is offering special training to its most promising junior salespeople. To qualify for the program, a sales associate must take verbal and quantitative exams, and earn a combined score:

$$C = 2\ (\text{Verbal Score}) + 3\ (\text{Quantitative Score})$$

of 550 points or more. Assume that verbal scores can be modeled using a normal distribution with mean 80 and standard deviation 10, quantitative scores can be modeled using a normal distribution with mean 120 and standard deviation 15, and scores are independent.

Find the probability that a junior salesperson qualifies for the special training program.

4.8 Moment Generating Functions

The moment generating function (MGF) of a random variable X is the function

$$M(t) = E(e^{tX}) = \begin{cases} \sum_{x\in\mathcal{R}} e^{tx}\ p(x) & \text{when } X \text{ is discrete} \\ \\ \int_{\mathcal{R}} e^{tx}\ f(x)\ dx & \text{when } X \text{ is continuous} \end{cases}$$

if the expectation exists, where t is a real number.

Moment generating functions have several useful properties, including:

Property A: If $M(t)$ exists for all t in an open interval containing 0, then it uniquely determines the probability distribution of X.

Property B: If $M(t)$ exists for all t in an open interval containing 0, then

$$M^{(k)}(0) = E(X^k) \quad \text{for } k = 0, 1, 2, \ldots,$$

where $M^{(k)}(t)$ is the k^{th} derivative of $M(t)$.

Idea of proof of Property B when $k = 1$: Let X be a discrete random variable with density function $p(x) = P(X = x)$ and moment generating function $M(t) = E(e^{tX})$.

Under suitable conditions on $M(t)$,

$$M'(0) = \frac{d}{dt}\left(\sum_{x\in\mathcal{R}} e^{tx}\ p(x)\right)\Bigg|_{t=0} = \left(\sum_{x\in\mathcal{R}} x\ e^{tx}\ p(x)\right)\Bigg|_{t=0} = \left(\sum_{x\in\mathcal{R}} x\ p(x)\right) = E(X).$$

Using these properties: Suppose that $M(t)$ exists for all t in an open interval containing 0.

Then we know from Property B that the Maclaurin series expansion of $M(t)$ must be

$$M(t) = \sum_{k=0}^{\infty} \frac{M^{(k)}(0)}{k!} t^k = \sum_{k=0}^{\infty} \frac{E(X^k)}{k!} t^k,$$

and we know from Property A that $M(t)$ uniquely defines the distribution.

Thus, the sequence of expected values

$$E(X^k) \quad \text{for } k = 0, 1, 2, \ldots$$

(known as the sequence of k^{th} *moments*) must define the distribution uniquely.

The next useful properties of moment generating functions are

Property C: Let $M_X(t)$ be the MGF of X, $M_Y(t)$ be the moment generating function of Y, and suppose that $Y = a + bX$. Then

$$M_Y(t) = e^{at} M_X(bt) \quad \text{for all } t.$$

Property D: If X and Y are independent random variables and $W = X + Y$,

$$M_W(t) = M_X(t) M_Y(t) \quad \text{for all } t.$$

More generally, if $W = \sum_{i=1}^{k} X_i$ is the sum of k mutually independent random variables, then $M_W(t) = \prod_{i=1}^{k} M_{X_i}(t)$.

Exercise. Use the definition of the moment generating function to verify Properties C and D.

Exercise. Let X be an exponential random variable with parameter $\lambda = 1$. Use the moment generating function of X to find $E(X^k)$ for $k = 1, 2, 3, 4$.

Exercise. Let Z be the standard normal random variable. Use the moment generating function of Z to find $E(Z^k)$ for $k = 1, 2, 3, 4$.

Exercise. (a) Demonstrate that the moment generating function of a Poisson random variable with parameter λ is $M(t) = e^{\lambda(e^t-1)}$.

(b) Let X be a Poisson random variable with parameter λ, let Y be a Poisson random variable with parameter μ, and let $W = X + Y$. If X and Y are independent, use moment generating functions to demonstrate that W is a Poisson random variable with parameter $\lambda + \mu$.

Exercise. Let X be a normal random variable with mean μ and standard deviation σ, and let Z be the standard normal random variable. Use the facts that

$$X = \mu + \sigma Z$$

and $M_Z(t) = e^{t^2/2}$ for all t to find a general form for the moment generating function of a normal random variable.

Exercise. Let X be a normal random variable with mean μ_x and standard deviation σ_x, Y be a normal random variable with mean μ_y and standard deviation σ_y, and assume that X and Y are independent.

Use moment generating functions to demonstrate that $W = a + bX + cY$ is a normal random variable, where a, b and c are constants. Identify the parameters of the distribution.