1 The Expectation of a Random Variable

Certain numbers associated with a random variable's distribution often provide a succinct way of summarizing the distribution. If only two numbers are used to describe a distribution, one is usually a measure of the *center* of the distribution and the second measures how *spread* out it is.

Numbers describing a distribution are commonly defined in terms of *expected values*.

1.1 Expectation of a Discrete RV

The expected value (or mean) of a discrete random variable X with pmf p_X is

$$E(X) = \mu_X = \sum_x x\, p_X(x),$$

where the sum extends over all x such that $p_X(x) > 0$.

Example 24 *Expectation Number of Spots on Two Dice.*

The number of spots on two dice has the following pmf:

x	2	3	4	5	6	7	8	9	10	11	12
$p_X(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

The expected number of spots is

$$\begin{aligned} E(X) &= 2\left(\frac{1}{36}\right) + 3\left(\frac{2}{36}\right) + \cdots + 12\left(\frac{1}{36}\right) \\ &= \frac{2}{36} + \frac{6}{36} + \cdots + \frac{12}{36} = \frac{252}{36} = 7 \end{aligned}$$

Example 12 Again

x	0	1	2	3	4
$P(X=x)$	0.5	0.25	0.125	0.0625	0.0625

Then

$$\begin{aligned} E(X) &= (0)(0.5)+(1)(0.25)+(2)(0.125)+(3)(0.0625)+(4)(0.0625) \\ &= 0.9375. \end{aligned}$$

Example 25 *Expectation of a Poisson Random Variable.* The rv X has a *Poisson* *(λ)* distribution if its pmf is

$$p_X(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, \ldots, \\ 0, & \text{otherwise,} \end{cases}$$

where λ is a positive constant. Find $E(X)$ when X has a Poisson (λ) distribution.

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!} \\ &= \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!} = \lambda \sum_{y=0}^{\infty} \frac{\lambda^y e^{-\lambda}}{y!} = \lambda. \end{aligned}$$

Remark: This example uses a common ”trick” in finding expectations:

Try to factor a constant, C, out of the sum (or integral) so that the “new” summand (or integrand) becomes a pmf or pdf. Then we know that the sum (or integral) is 1, and the expectation is C.

Remark: An important observation is that *expectations need not exist.*

When X is discrete and takes on an infinite number of values, then the sum $\sum_x x\, p_X(x)$ may not exist or may not be finite.

Of course, if X takes on only finitely many values, then $E(X)$ does exist and is finite.

Example 31 *Expectation of a Binomial Random Variable* A rv X has a *binomial (n,θ)* distribution if its pmf is

$$p_X(x) = \begin{cases} \binom{n}{x}\theta^x(1-\theta)^{n-x}, & x = 0, 1, \ldots, n, \\ 0, & \text{otherwise,} \end{cases}$$

where $0 < \theta < 1$. Find $E(X)$ when X has a binomial (n,θ) distribution.

$$\begin{aligned} E(X) &= \sum_{x=0}^{n} x \cdot \binom{n}{x}\theta^x(1-\theta)^{n-x} = \sum_{x=1}^{n} x \cdot \frac{n!}{x!(n-x)!}\theta^x(1-\theta)^{n-x} \\ &= n\theta \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!}\theta^{x-1}(1-\theta)^{n-x}. \end{aligned}$$

Make the change of variable $y = x - 1$ in the last sum to get

$$E(X) = n\theta \sum_{y=0}^{n-1} \frac{(n-1)!}{y!(n-1-y)!}\theta^y(1-\theta)^{n-1-y} = n\theta.$$

The last equality holds since we are summing the pmf of a binomial $(n-1, \theta)$ rv.

1.2 Expectation of a Continuous RV

When X is continuous with pdf f_X, the expected value (or mean) of X is

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x\, f_X(x)\, dx.$$

When X is continuous and its density is positive only over an interval of finite length, then $E(X)$ exists and is finite. Otherwise, there are cases where $E(X)$ either doesn't exist or isn't finite.

Important intuition:

In both the discrete and continuous cases, the expected value of X (when it exists and is finite) has the interpretation that it is the *average value of* X in a large number of repetitions of the experiment.

Example 26 *Expectation of an uniform random variable.* Suppose X has the uniform$[L, R]$ distribution with pdf

$$f_X(x) = \frac{1}{R-L} I_{[L,R]}(x).$$

Then

$$E(X) = \int_L^R x \frac{1}{R-L} dx = \frac{1}{R-L} \frac{x^2}{2}\bigg|_L^R = \frac{R^2 - L^2}{2(R-L)} = \frac{L+R}{2}.$$

Example 27 *Expectation of an exponential random variable.* Suppose X has the exponential density

$$f_X(x) = \frac{1}{\lambda} e^{-x/\lambda} I_{(0,\infty)}.$$

Find the expected value of X.

$$\begin{aligned} E(X) &= \int_0^\infty x \cdot \frac{1}{\lambda} e^{-x/\lambda}\, dx \\ &= \int_0^\infty (\lambda y) \cdot \frac{1}{\lambda} e^{-y}\, d(\lambda y) \\ &= \lambda \int_0^\infty y e^{-y}\, dy \\ &= \lambda, \end{aligned}$$

where the last step follows from Example 23 in Chapter 2. Why?

So, we have again used the trick described in Example 25.

1.3 Expectations of Functions of X

Consider a random variable $Y = g(X)$, and let p_Y or f_Y be the pmf or pdf of Y, respectively. According to the previous definition,

$$E(Y) = \begin{cases} \sum_y y\, p_Y(y), & \text{if } Y \text{ is discrete,} \\ \\ \int_{-\infty}^{\infty} y\, f_Y(y)\, dy, & \text{if } Y \text{ is continuous.} \end{cases}$$

One can prove (p. 134 and 144 of the text) that $E[g(X)]$ can also be computed as follows:

$$E[g(X)] = \begin{cases} \sum_x g(x)\, p_X(x), & X \text{ discrete,} \\ \\ \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx, & X \text{ continuous.} \end{cases}$$

We may also be interested in the expectation of a function of several random variables. Consider the random variables $X_1, \ldots, X_n$ and the function $Y = g(X_1, \ldots, X_n)$. If $X_1, \ldots, X_n$ are continuous random variables with joint pdf f, then

$$E(Y) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n.$$

If $X_1, \ldots, X_n$ are discrete random variables with joint pmf p, then

$$E(Y) = \sum_{x_1} \cdots \sum_{x_n} g(x_1, \ldots, x_n)\, p(x_1, \ldots, x_n).$$

Example 27 Expectation of a Function of a Poisson RV

Let X have the Poisson distribution given in Example 25. Find the expected value of $g(X) = 2^X$.

$$
\begin{aligned}
E(2^X) &= \sum_{x=0}^{\infty} 2^x \cdot \frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=0}^{\infty} \frac{(2\lambda)^x e^{-\lambda}}{x!} \\
&= \frac{e^{-\lambda}}{e^{-2\lambda}} \sum_{x=0}^{\infty} \frac{(2\lambda)^x e^{-2\lambda}}{x!} = e^{\lambda} \cdot 1 = e^{\lambda}.
\end{aligned}
$$

Example 28 Expectation of a Function of a Bivariate RV

Let (X, Y) have the joint pdf

$$f(x,y) = 2,\ 0 < x < 1,\ 0 < y < 1, 0 < x + y < 1.$$

Find the expected value of $g(X, Y) = XY$.

$$\begin{aligned}
E(XY) &= \int_0^1 \int_0^{1-x} xy\, 2\, dy\, dx = \int_0^1 2x \frac{y^2}{2}\Big|_0^{1-x} dx \\
&= \int_0^1 x(1-x)^2 dx = \int_0^1 (x - 2x^2 + x^3) dx \\
&= \left(\frac{x^2}{2} - 2\frac{x^3}{3} + \frac{x^4}{4} \right)\Big|_0^1 = \frac{1}{2} - \frac{2}{3} + \frac{1}{4} = \frac{1}{12}
\end{aligned}$$

2 Properties of expectations

2.1 Expectation of a linear function of X

Let $Y = aX + b$, where a and b are constants. If $E(X)$ exists, then

$$E(Y) = aE(X) + b.$$

Proof: Consider the continuous case.

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} (ax+b) f_X(x)\, dx \\ &= \int_{-\infty}^{\infty} ax f_X(x)\, dx + \int_{-\infty}^{\infty} b \cdot f_X(x)\, dx \\ &= a \int_{-\infty}^{\infty} x f_X(x)\, dx + b \int_{-\infty}^{\infty} f_X(x)\, dx \\ &= aE(X) + b. \end{aligned}$$

Remark: If g(x) is a nonlinear function, $E(g(X)) \neq g(E(X))$ in most situations.

Example: In Example 25, we let X have a Poisson(λ) distribution and considered $g(X) = 2^X$. We earlier showed that $E(X) = \lambda$. Thus,

$$E(g(X)) = E[2^X] = e^\lambda \neq 2^\lambda = g(E(X)).$$

2.2 Expectation of a sum of random variables

Let $Y = a + b_1X_1 + b_2X_2 + \cdots + b_nX_n$ and assume that $E(X_i)$ exists, $i = 1, \ldots, n$. Then

$$E(Y) = a + b_1E(X_1) + \cdots + b_nE(X_n).$$

This result is easily proven in the continuous case using the fact that the integral of a sum is the sum of integrals.

2.3 Expectation of a product of independent random variables

Suppose that $X_1, \ldots, X_n$ are independent random variables, and let $h_1, \ldots, h_n$ be functions such that $E[h_i(X_i)]$ exists, $i = 1, \ldots, n$. Then

$$E\left[\prod_{i=1}^{n} h_i(X_i)\right] = \prod_{i=1}^{n} E[h_i(X_i)].$$

We'll prove this in the case where X_1, X_2 are continuous with joint pdf f.

Let f_i be the marginal pdf of X_i, $i = 1, 2$. Then by definition of independence,

$$f(x_1, x_2) = f_1(x_1) \times f_2(x_2).$$

By definition of expectation,

$$\begin{aligned}
E\left[h_1(X_1)h_2(X_2)\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(x_1)h_2(x_2)f(x_1,x_2)\,dx_1dx_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(x_1)h_2(x_2)f_1(x_1)f_2(x_2)dx_1dx_2 \\
&= \int_{-\infty}^{\infty} h_1(x_1)f_1(x_1)dx_1 \times \int_{-\infty}^{\infty} h_2(x_2)f_2(x_2)dx_2 \\
&= E\left[h_1(X_1)\right]E\left[h_2(X_2)\right].
\end{aligned}$$

Example 29 Let X_1 and X_2 be independent and identically distributed (i.i.d.) random variables, with each having a uniform distribution on the interval (L, R), where $L > 0$. Determine $E(X_1/X_2^2)$.

Since X_1 and X_2 are independent,

$$\begin{aligned} E(X_1/X_2^2) &= E(X_1)E(X_2^{-2}) \\ &= \frac{(L+R)}{2} \cdot E(X_2^{-2}). \end{aligned}$$

Now,

$$\begin{aligned} E(X_2^{-2}) &= \int_L^R x^{-2}(R-L)^{-1}\,dx \\ &= \left.\frac{-x^{-1}}{(R-L)}\right|_L^R = \frac{1}{LR}. \end{aligned}$$

Hence, $E(X_1/X_2^2) = (L+R)/(2LR)$.

2.4 Monotonicity of Expectation

Let X be a random variable where $P[X \geq 0] = 1$. Then

$$E(X) \geq 0.$$

We will show this in the discrete case. Let $R_X = \{x_1, x_2, ...\}$ be the set of values where $P[X = x] > 0$. Then $x_i \geq 0$ by assumption and

$$E(X) = \sum_{x \in R_X} x P[X = x] \geq 0$$

since each term in the sum is nonnegative.

Let X and Y be random variables where $X \leq Y$. Then

$$E(X) \leq E(Y).$$

Note: $X \leq Y$ means $X(s) \leq Y(s)$ for all $s \in \mathcal{S}$.

3 Variance of a Random Variable

Suppose X is a random variable such that $E(X) = \mu_X$ exists. The *variance* of X (if it exists) is defined to be

$$\sigma_X^2 = \mathrm{Var}(X) = E[(X - \mu_X)^2].$$

The variance provides a simple way of summarizing the amount of *variability* or *dispersion* in the distribution of a rv. It is particularly nice for comparing two or more distributions.

The *standard deviation* of a random variable is defined by

$$\sigma_X = \sqrt{\mathrm{Var}(X)} = \sqrt{E[(X - \mu_X)^2]}.$$

The standard deviation of a random variable X is more often reported than the variance since standard deviation has the same units as X.

Example 24 again *Variance of the Number of Spots on Two Dice.*

The number of spots on two dice has the following pmf:

x	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

The variance and standard deviation of the number of spots are

$$\begin{aligned} E((X-\mu)^2) &= (2-7)^2\left(\frac{1}{36}\right)+(3-7)^2\left(\frac{2}{36}\right)+\cdots+(12-7)^2\left(\frac{1}{36}\right) \\ &= \frac{25}{36}+\frac{32}{36}+\cdots+\frac{25}{36}=\frac{210}{36}=\frac{35}{6}, \\ \sigma_X &= \sqrt{\frac{35}{6}}. \end{aligned}$$

Another Way of Finding the Variance

Let $Y = (X-7)^2$. Then $\sigma^2 = E[(X-7)^2] = E(Y)$.

We can use the pmf of Y to find $E(Y)$. Using the fact that

$$p_Y(y) = P[Y=y] = P[(X-7)^2 = y] = \sum_{\{x:(x-7)^2=y\}} p_X(x),$$

the pmf of Y is

y	0	1	4	9	16	25
$g(y)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

.

Then

$$\begin{aligned} E(Y) &= 0\left(\frac{6}{36}\right) + 1\left(\frac{10}{36}\right) + 4\left(\frac{8}{36}\right) + 9\left(\frac{6}{36}\right) + 16\left(\frac{4}{36}\right) + 25\left(\frac{2}{36}\right) \\ &= \frac{210}{36} = \frac{35}{6}. \end{aligned}$$

Example 26 Again Find the variance of a random variable X that is uniformly distributed over (L, R). We know that $\mu = (L+R)/2$.

$$\begin{aligned}
\mathrm{Var}(X) &= \int_L^R (x-\mu)^2 (R-L)^{-1}\, dx \\
&= (R-L)^{-1} \int_L^R (x-\mu)^2 \, d(x-\mu) \\
&= \frac{1}{3(R-L)} (x-\mu)^3 \Big|_L^R \\
&= (R-L)^2/12.
\end{aligned}$$

The standard deviation of a uniform (R, L) rv is

$$\sigma = \frac{R-L}{2\sqrt{3}}.$$

3.1 Properties of Variance

1. The variance of X is equal to

$$E(X^2) - [E(X)]^2 = E(X^2) - \mu_X^2.$$

Proof: Since $(X - \mu_X)^2 = X^2 - 2X\mu_X + \mu_X^2$,

$$\begin{aligned} \mathrm{Var}(X) &= E[(X - \mu_X)^2] \\ &= E(X^2) - 2\mu_X E(X) + \mu_X^2 \\ &= E(X^2) - \mu_X^2. \end{aligned}$$

2. $\mathrm{Var}(X) = 0$ if and only if there is a constant c such that $P(X = c) = 1$.

3. If a and b are constants, then $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$.

   Proof: We know that $E(aX+b) = aE(X)+b = a\mu_X + b$, and so

$$\begin{aligned} \mathrm{Var}(aX+b) &= E\{[(aX+b)-(a\mu_X+b)]^2\} \\ &= E\{[a(X-\mu_X)]^2\} \\ &= E\{a^2(X-\mu_X)^2\} \\ &= a^2E\{(X-\mu_X)^2\} \\ &= a^2\,\mathrm{Var}(X). \end{aligned}$$

4. Suppose that $X_1,\ldots,X_n$ are independent random variables, and let $a_1,\ldots,a_n$ be constants. Then

$$\mathrm{Var}(a_1X_1+\cdots+a_nX_n) = \sum_{i=1}^{n} a_i^2\,\mathrm{Var}(X_i).$$

Remarks on properties of variance:

- Take $a = 1$ in Property 3. Then we see that $\mathrm{Var}(X + b) = \mathrm{Var}(X)$. This means that *shifting* a distribution to the left or right has no effect on its variance.

- Property 3 implies that the standard deviation of $aX + b$ is $|a|\sigma$, where σ is the standard deviation of X.

- If we take $a_1 = \cdots = a_n = 1$ in Property 4, the result becomes

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

  So, the variance of a sum of *independent* random variables is the sum of variances. This result is **not** necessarily true when $X_1, \ldots, X_n$ are not independent.

Example 31 *Variance of a binomial random variable.* Suppose X has the binomial distribution. Find $\mathrm{Var}(X)$.

We'll use the fact that $\mathrm{Var}(X) = E(X^2) - [E(X)]^2$. We know that $E(X) = n\theta$. So we need to find $E(X^2)$.

Suppose we could find $E[X(X-1)] = E(X^2) - E(X) = E(X^2) - n\theta$. Then we have $E(X^2) = E[X(X-1)] + n\theta$.

Assume $n \geq 2$. Now,

$$\begin{aligned} E[X(X-1)] &= \sum_{x=0}^{n} x(x-1)\binom{n}{x}\theta^x(1-\theta)^{n-x} \\ &= \sum_{x=2}^{n} x(x-1)\binom{n}{x}\theta^x(1-\theta)^{n-x} \\ &= n(n-1)\sum_{x=2}^{n}\binom{n-2}{x-2}\theta^x(1-\theta)^{n-x}. \end{aligned}$$

Now, make the change of variable $y = x - 2$ in the last sum. This gives

$$\begin{aligned} E[X(X-1)] &= n(n-1)\sum_{x=2}^{n}\binom{n-2}{x-2}\theta^{x}(1-\theta)^{n-x} \\ &= n(n-1)\theta^{2}\sum_{y=0}^{n-2}\binom{n-2}{y}\theta^{y}(1-\theta)^{n-2-y}. \end{aligned}$$

The last sum equals 1. Why? So, we have $E[X(X-1)] = n(n-1)\theta^2$, which implies that

$$\begin{aligned} E(X^2) &= n(n-1)\theta^2 + n\theta \\ &= (n\theta)^2 + n\theta(1-\theta), \end{aligned}$$

and hence

$$\mathrm{Var}(X) = n\theta(1-\theta).$$

4 Covariance and Correlation

In practice it is often of interest to know how two variables are related. When X increases or decreases, how does Y behave?

When the relationship between X and Y is relatively simple, the *covariance* and/or *correlation* are good measures for summarizing the relationship.

The covariance between X and Y is denoted $\mathrm{Cov}(X, Y)$ and defined by

$$\mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)],$$

where $\mu_X = E(X)$ and $\mu_Y = E(Y)$.

$\mathrm{Cov}(X, Y)$ measures the tendency of X and Y to be on the same (or opposite) sides of their respective means.

The correlation between X and Y is a scaled version of the covariance. It (correlation) does not depend on the measurement units of X and Y.

The correlation is denoted $\mathrm{Corr}(X, Y)$ and defined by

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y},$$

where σ_X and σ_Y are the standard deviations of X and Y, respectively. We can also use $\rho,\ \rho(X, Y)$, or $\rho_{X,Y}$ as notation for $\mathrm{Corr}(X, Y)$.

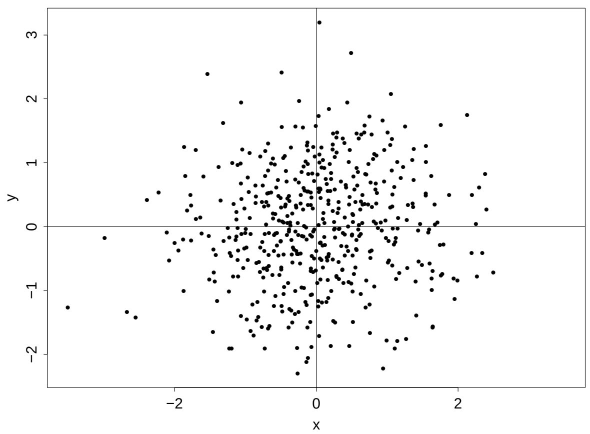
An important property of correlation is that

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

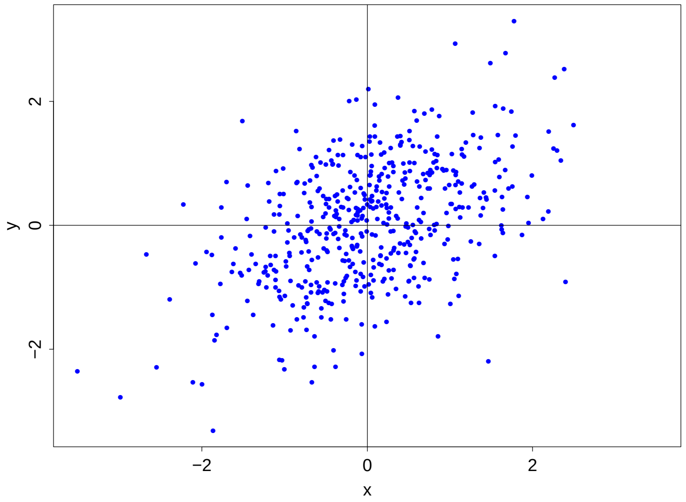
See the proof on p. 186 of your text.

When X and Y are on the *same* side of their respective means with very high probability, then $\mathrm{Corr}(X, Y)$ will be close to 1. When X and Y are on *opposite* sides of their respective means with very high probability, $\mathrm{Corr}(X, Y)$ is close to -1.

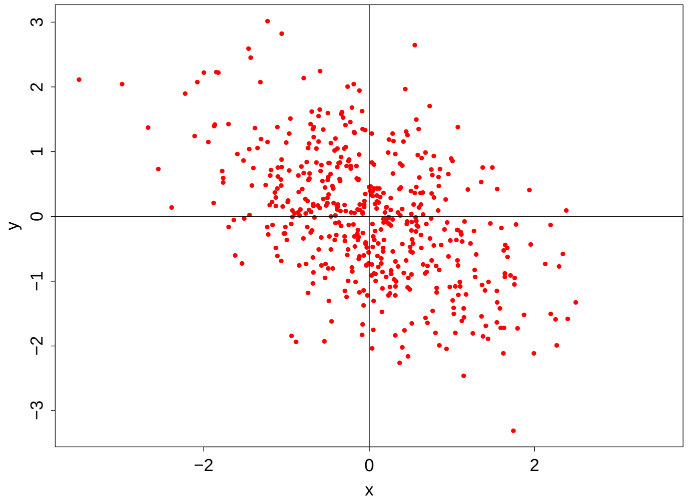
The last comments are best illustrated graphically. Suppose we repeat the experiment that generates X and Y hundreds of times. Each time the experiment is repeated the result is an (x, y) pair. We could plot the hundreds of (x, y) pairs on a *scatter plot*.



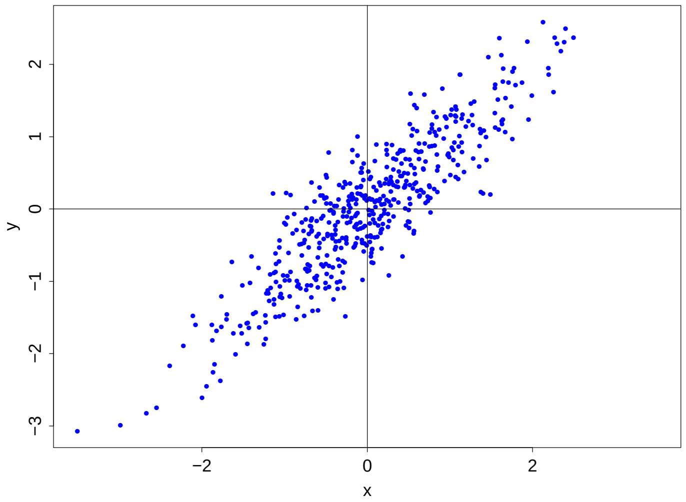
$$\rho(X, Y) = 0.0$$



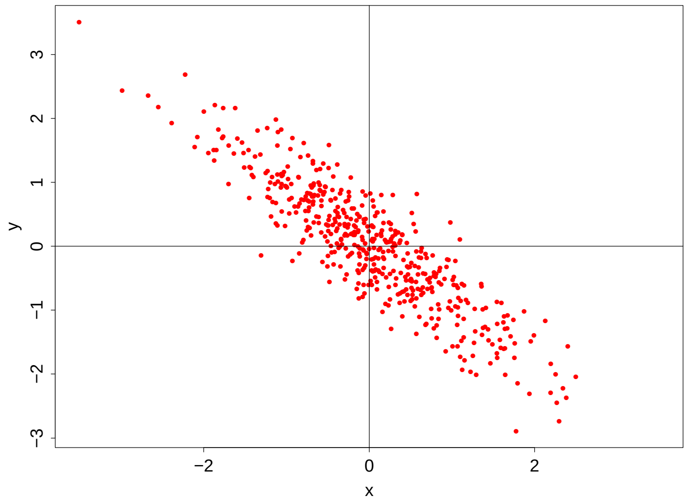
$$\rho(X, Y) = 0.5$$



$$\rho(X, Y) = -0.5$$



$$\rho(X, Y) = 0.9$$



$$\rho(X, Y) = -0.9$$

4.1 Properties of Covariance and Correlation

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

  Proof: We expand the product in the definition of covariance:

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\
&= E[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y] \\
&= E[XY] - \mu_X E(Y) - \mu_Y E(X) + \mu_X\mu_Y \\
&= E[XY] - E(X)E(Y)
\end{aligned}$$

- If X and Y are independent, $0 < \sigma_X < \infty$ and $0 < \sigma_Y < \infty$, then

$$\mathrm{Cov}(X,Y) = 0 = \mathrm{Corr}(X,Y).$$

- $\mathrm{Cov}(X,X) = E[(X-\mu_X)(X-\mu_X)] = \mathrm{Var}(X)$.

- Suppose a and b are constants with $a \neq 0$ and that $0 < \sigma_X < \infty$. Then if $Y = aX + b$,

$$\mathrm{Corr}(X, Y) = \begin{cases} 1, & a > 0, \\ -1, & a < 0. \end{cases}$$

Proof: We know that $\mu_Y = a\mu_X + b$, and hence

$$\begin{aligned} \mathrm{Cov}(X, Y) &= E[(X - \mu_X)(aX - a\mu_X)] \\ &= a\mathrm{Var}(X). \end{aligned}$$

We also know that $\sigma_Y = |a|\sigma_X$, and hence

$$\mathrm{Corr}(X, Y) = \frac{a}{|a|} = \pm 1,$$

depending on the sign of a.

Note: $\text{Cov}(X, Y) = 0$ is a weaker condition than independence of X and Y.

- "X and Y independent" $\Rightarrow \text{Cov}(X, Y) = 0$.
- Converse of previous implication is not true, i.e., there are cases where $\text{Cov}(X, Y) = 0$, but X and Y are **not** independent.

Example: Suppose (X, Y) are discrete rvs with joint pmf

$$p_{X,Y}(x, y) = \begin{cases} \frac{1}{4}, & (x, y) \in \{(0, 1), (1, 0), (0, -1), (-1, 0), \\ 0, & \text{otherwise.} \end{cases}$$

First, $E(X) = E(Y) = 0$. Then

$$\text{Cov}(X, Y) = E[XY] = \frac{1}{4}[(0)(1)+(1)(0)+(0)(-1)+(-1)(0)] = 0.$$

Example 12 again Find the covariance and correlation of X and Y in the coin tossing example.

		y					
		0	1	2	3	4	$p_X(x)$
	0	1/16	3/16	3/16	1/16	0	8/16
	1	0	1/16	2/16	1/16	0	4/16
x	2	0	0	1/16	1/16	0	2/16
	3	0	0	0	1/16	0	1/16
	4	0	0	0	0	1/16	1/16
	$p_Y(y)$	1/16	4/16	6/16	4/16	1/16	

Previously we found that $E(X) = 15/16$ and $E(Y) = 4(0.5) = 2$. Next

$$E(XY) = (1)(1)\left(\frac{1}{16}\right) + (1)(2)\left(\frac{2}{16}\right) + \cdots + (4)(4)\left(\frac{1}{16}\right) = \frac{43}{16}$$

$$\mathrm{Cov}(X, Y) = \frac{43}{16} - \left(\frac{15}{16}\right)(2) = \frac{13}{16}.$$

We need to calculate the variances of X and Y:

$$E(X^2) = (1^2)\left(\frac{4}{16}\right) + (2^2)\left(\frac{2}{16}\right) + (3^2 + 4^2)\left(\frac{1}{16}\right) = \frac{37}{16}$$

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = \frac{37}{16} - \left(\frac{15}{16}\right)^2 = \frac{367}{256}.$$

$$\mathrm{Var}(Y) = 4(0.5)(1 - 0.5) = 1$$

Then the correlation of X and Y is

$$\mathrm{Corr}(X, Y) = \frac{\frac{13}{16}}{\sqrt{\left(\frac{367}{256}\right)(1)}} = 0.4798$$

Example 19 again Let X and Y have joint pdf

$$f(x,y) = \begin{cases} 3(x+y), & 0 < x < 1, 0 < y < 1, \\ & 0 < x+y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Earlier we found that the marginal pdf of X was

$$f_1(x) = \frac{3}{2}(1-x^2),\ 0 < x < 1.$$

The first two moments are

$$E(X) = \frac{3}{8} \text{ and } \mathrm{Var}(X) = \frac{19}{320}.$$

The covariance of X and Y is

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E(XY)-\mu_X\mu_Y=\int_0^1\int_0^{1-x} xy\,3(x+y)dydx-\left(\frac{3}{8}\right)^2\\ &= \frac{1}{10}-\left(\frac{3}{8}\right)^2=-\frac{13}{320}\end{aligned}$$

The correlation of X and Y is

$$\mathrm{Corr}(X,Y)=\frac{-\frac{13}{320}}{\sqrt{\frac{19}{320}}\sqrt{\frac{19}{320}}}=-\frac{13}{19}=-0.6842$$

Example 28 again Let X and Y have joint pdf

$$f(x,y) = \begin{cases} 2, & 0 < x < 1, 0 < y < 1, \\ & 0 < x + y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

The marginal pdf of X is

$$f_1(x) = \int_0^{1-x} 2dy = 2(1-x),\ 0 < x < 1.$$

The first two moments are

$$E(X) = \frac{1}{3} \text{ and } \mathrm{Var}(X) = \frac{1}{18}.$$

The covariance of X and Y is

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E(XY)-\mu_X\mu_Y = \int_0^1\int_0^{1-x} xy\,2\,dydx - \left(\frac{1}{3}\right)^2 \\ &= \frac{1}{12}-\left(\frac{1}{3}\right)^2 = -\frac{1}{36}\end{aligned}$$

The correlation of X and Y is

$$\mathrm{Corr}(X,Y) = \frac{-\frac{1}{36}}{\sqrt{\frac{1}{18}}\sqrt{\frac{1}{18}}} = -\frac{1}{2}$$

4.2 Variance of a linear combination

Suppose X and Y are jointly distributed random variables and define $Z = aX + bY$, where a and b are constants. What is $\text{Var}(Z)$?

We know that $E(Z) = a\mu_X + b\mu_Y$, and so

$$\begin{aligned}
\text{Var}(Z) &= E\left[(aX + bY - a\mu_X - b\mu_Y)^2\right] \\
&= E\left[(aX - a\mu_X + bY - b\mu_Y)^2\right] \\
&= E\big[a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 + 2ab(X - \mu_X)(Y - \mu_Y)\big] \\
&= a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\,\text{Cov}(X, Y).
\end{aligned}$$

What if $a = b = 1$ and $\text{Cov}(X, Y) = 0$? Then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

So, the variance of a sum is the sum of variances if and only if $\text{Cov}(X, Y) = 0$.

- We extend the result to a linear combination of several random variables:

  Suppose $U = a + \sum_{i=1}^{n} b_i X_i$. Then

$$\begin{aligned}\mathrm{Var}(U) &= \sum_{i=1}^{n}\sum_{j=1}^{n} b_i b_j \mathrm{Cov}(X_i, X_j) \\ &= \sum_{i=1}^{n} b_i^2 \mathrm{Var}(X_i) + 2\sum_{i<j} b_i b_j \mathrm{Cov}(X_i, X_j).\end{aligned}$$

- This result can also be extended to the covariance of two linear combinations of random variables:

  Suppose $U = a + \sum_{i=1}^{n} b_i X_i$ and $V = c + \sum_{j=1}^{m} d_j Y_j$. Then

$$\mathrm{Cov}(U, V) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_i d_j \mathrm{Cov}(X_i, Y_j).$$

5 Moments and the Moment-Generating Function

The *moments* of a random variable X are (when they exist) the expectations of powers of X:

$$E(X^k), \quad k = 1, 2, \ldots .$$

Moments can be useful in describing the distribution of a rv. In certain situations, the set of all moments uniquely determines the distribution. (More on this shortly.)

Theorem: If $E(X^k)$ exists for some k, then $E(X^j)$ exists for $j = 1, 2, \ldots, k-1$.

5.1 Central Moments

Assuming existence of the moments (as defined before), the *central moments* of X are

$$E[(X-\mu)^k], \quad k = 1, 2, \ldots .$$

Remark: When the moments exist, the central moments and the moments are functions of each other. For example, the second central moment can be expressed as

$$\begin{aligned} E[(X-\mu)^2] &= E(X^2) - \mu^2 \\ \\ E[(X-\mu)^3] &= E(X^3) - 3\mu E(X^2) + 3\mu^2 E(X) - \mu^3 \\ &= E(X^3) - 3\mu E(X^2) + 2\mu^3 \end{aligned}$$

5.2 Moment generating function

Consider the following function of s:

$$M_X(s) = E[e^{sX}].$$

If there is a positive number s_0 such that the last expectation exists for all $|s| < s_0$, then $M_X(s)$, $|s| < s_0$, is called the *moment generating function*, or mgf, of X.

When the mgf of X exists, then all the moments of X exist and are finite.

Furthermore, we may find the moments of X from M_X in the following way:

$$E(X^k) = \left. \frac{d^k M_X(s)}{ds^k} \right|_{s=0}.$$

For example, $M'_X(0) = E(X)$ and $M''_X(0) = E(X^2)$.

Example 12 Again Suppose that X has the pmf

x	0	1	2	3	4
$P(X = x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$

Then the moment generating function of X is

$$M_X(s) = E(e^{sX}) = \frac{1}{2}e^0 + \frac{1}{4}e^s + \frac{1}{8}e^{2s} + \frac{1}{16}(e^{3s} + e^{4s})$$

We next take the derivative with respect to s:

$$M'_X(s) = \frac{dM_X(s)}{ds} = \frac{1}{4}e^s + \frac{1}{8}2e^{2s} + \frac{1}{16}(3e^{3s} + 4e^{4s}).$$

Evaluate this at $s = 0$ to obtain the mean of X:

$$M'_X(0) = \frac{1}{4} + \frac{1}{8}(2) + \frac{1}{16}(3 + 4) = \frac{15}{16}.$$

Example 32 *Mgf of binomial distribution.* Suppose that X has the binomial distribution with pmf

$$p_X(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \; x = 0, 1, \ldots, n.$$

The mgf of X is given by

$$\begin{aligned} M_X(s) &= E(e^{sX}) = \sum_{x=0}^{n} e^{sx} \binom{n}{x} \theta^x (1-\theta)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (e^s \theta)^x (1-\theta)^{n-x} \\ &= (\theta e^s + 1 - \theta)^n \end{aligned}$$

We can find the moments of the binomial rv X by differentiation:

$$E(X) = M'(0) = \frac{d}{ds}(\theta e^s + 1 - \theta)^n \Big|_{s=0} = n(\theta e^s + 1 - \theta)^{n-1}\theta e^s \Big|_{s=0} = n\theta.$$

$$\begin{aligned} E(X^2) &= M''(0) = \frac{d^2}{ds^2}(\theta e^s + 1 - \theta)^n \Big|_{s=0} \\ &= \frac{d}{ds}\left[n(\theta e^s + 1 - \theta)^{n-1}\theta e^s\right] \Big|_{s=0} \\ &= \left[n(n-1)(\theta e^s + 1 - \theta)^{n-2}\theta^2 e^{2s} + n(\theta e^s + 1 - \theta)^{n-1}\theta e^s\right] \Big|_{s=0} \\ &= n(n-1)\theta^2 + n\theta \end{aligned}$$

The variance of X is

$$\mathrm{Var}(X) = n(n-1)\theta^2 + n\theta - (n\theta)^2 = n\theta - n\theta^2 = n\theta(1-\theta)$$

Example 33 *Mgf of normal distribution.* Let X have the normal distribution with mean μ_X and variance σ_X^2, i.e., its pdf is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left[-\frac{1}{2\sigma_X^2}(x-\mu_X)^2\right].$$

Find the mgf of X. We have

$$M_X(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x)\, dx.$$

First we'll make the change of variable $y = (x - \mu_X)/\sigma_X$.

The integral is then

$$\int_{-\infty}^{\infty} e^{s(\sigma_X y + \mu_X)} \cdot \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy = e^{s\mu_X} \int_{-\infty}^{\infty} e^{s\sigma_X y} \cdot \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy.$$

From this point, the trick is to *complete the square* and produce an integrand that is proportional to a normal density. Consider

$$\begin{aligned} e^{s\sigma_X y - \frac{1}{2}y^2} &= \exp\left[-\frac{1}{2}(y^2 - 2s\sigma_X y)\right] \\ &= \exp\left[-\frac{1}{2}(y^2 - 2s\sigma_X y + s^2\sigma_X^2 - s^2\sigma_X^2)\right] \\ &= \exp\left[\frac{(s\sigma_X)^2}{2}\right]\exp\left[-\frac{1}{2}(y - s\sigma_X)^2\right]. \end{aligned}$$

So now we have

$$\begin{aligned} M_X(s) &= e^{s\mu_X + s^2\sigma_X^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-(y-s\sigma_X)^2/2}\,dy \\ &= \exp\left(s\mu_X + \frac{s^2\sigma_X^2}{2}\right). \end{aligned}$$

How did we get the very last step?

Now, let's use this mgf to find the first two moments of a normal distribution. Our notation suggests that $\mu_X = E(X)$ and $\sigma_X^2 = \mathrm{Var}(X)$. Is this true?

$$
\begin{aligned}
M_X'(s) &= \frac{d\exp\left(s\mu_X + s^2\sigma_X^2/2\right)}{ds} \\
&= \exp\left(s\mu_X + s^2\sigma_X^2/2\right)\frac{d\left(s\mu_X + s^2\sigma_X^2/2\right)}{ds} \\
&= M_X(s)\left(\mu_X + s\sigma_X^2\right).
\end{aligned}
$$

So, $E(X) = M_X'(0) = \mu_X$.

To find the second moment, we compute

$$
M_X''(s) = \sigma_X^2 M_X(s) + \left(\mu_X + s\sigma_X^2\right) M_X'(s),
$$

and find $E(X^2) = M_X''(0) = \sigma_X^2 + \mu_X^2$. Therefore,

$$
\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = (\sigma_X^2 + \mu_X^2) - \mu_X^2 = \sigma_X^2.
$$

5.3 Properties of mgfs

1. $M_X(0) = 1$

2. Let a and b be constants, and M_X be the mgf of X. Then the mgf of $Y = aX + b$ is

$$M_Y(s) = E(e^{sY}) = e^{bs} M_X(as).$$

3. Let $X_1, \ldots, X_n$ be independent random variables with respective mgfs $M_1, \ldots, M_n$. Then the mgf, M, of $X_1 + \cdots + X_n$ is

$$M(s) = \prod_{i=1}^{n} M_i(s).$$

4. Suppose the mgfs of the random variables X and Y exist, and call them M_X and M_Y, respectively. Then the distribution of X is the same as that of Y if and only if M_X is the same as M_Y.

A corollary to property 4 is that when the mgfs of X and Y exist, then X and Y have the same distribution if and only if the moments of X equal the corresponding moments of Y.

Here's a very interesting fact, however. There exist cases where

$$E(X^k) = E(Y^k), \quad k = 1, 2, \ldots,$$

and yet X and Y have remarkably different distributions.

Example 34 *Two different distributions that have all the same moments.* Consider the two pdfs

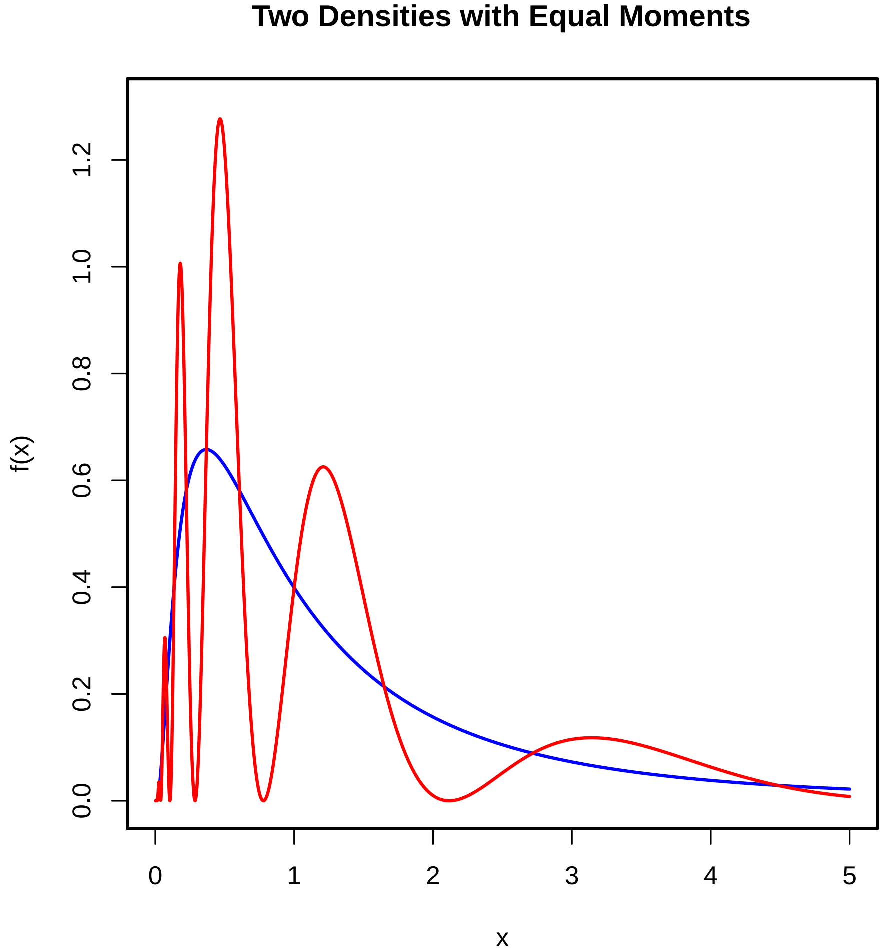
$$f_1(x) = \frac{1}{\sqrt{2\pi x}} e^{-(\log x)^2/2} I_{(0,\infty)}(x)$$

and

$$f_2(x) = f_1(x)[1 + \sin(2\pi \log x)].$$

One can show that the moments exist for these two pdfs and, for $k = 1, 2, \ldots$,

$$\int_0^\infty x^k f_1(x)\, dx = \int_0^\infty x^k f_2(x)\, dx.$$



Blue – f_1 Red – f_2

The reason this example does not contradict property 4 (p. 132) is that the mgfs of f_1 and f_2 do not exist.

Combining properties 3 and 4 yields a powerful method for finding the distribution of a sum of independent random variables.

Given independent random variables $X_1, \ldots, X_n$ with mgfs $M_1, \ldots, M_n$, try to find the distribution of $X_1 + \cdots + X_n$ as follows:

- Find the mgf M of the sum using property 3.
- Since the mgf uniquely determines the distribution (by property 4), if we recognize M as being the mgf of a known distribution, then we've found the distribution of $X_1 + \cdots + X_n$.

Example 35 *Distribution of a sum of Bernoulli rvs*

Let $X_1, \ldots, X_n$ be independent random variables such that

$$P(X_i = x) = \begin{cases} \theta, & x = 1 \\ 1-\theta, & x = 0 \\ 0, & \text{otherwise} \end{cases}$$

Each X_i has mgf $M_i(s) = E(e^{sX_i}) = (1-\theta)e^0 + \theta e^s = 1 - \theta + \theta e^s$.
Then the mgf of $Y = X_1 + \cdots + X_n$ is

$$\begin{aligned} M_Y(s) &= E(e^{Ys}) = \prod_{i=1}^{n} E(e^{X_i s}) \\ &= (1 - \theta + \theta e^s)^n \end{aligned}$$

We recognize this as the mgf of a binomial rv with n trials and probability of success θ.

Example 36 *Distribution of a sum of independent normal rvs*

Let $X_1, \ldots, X_n$ be independent random variables such that

$$X_i \sim N(\mu_i, \sigma_i^2), \quad i = 1, \ldots, n.$$

Using property 3 and the normal mgf, the mgf of $Y = X_1 + \cdots + X_n$ is

$$\begin{aligned} M_Y(s) &= \prod_{i=1}^{n} \exp\left(s\mu_i + s^2\sigma_i^2/2\right) \\ &= \exp\left(s\sum_{i=1}^{n}\mu_i + s^2\sum_{i=1}^{n}\sigma_i^2/2\right). \end{aligned}$$

From property 4 and Example 33, it immediately follows that

$$Y \sim N\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i^2\right).$$

6 Conditional Expectation and Prediction

Very simply, a *conditional expectation* is an ordinary expectation but defined with respect to a conditional distribution.

Suppose the conditional pmf or pdf of Y given $X = x$ is $p_{Y|X}(y|x)$ or $f_{Y|X}(y|x)$, respectively. Then

$$E[h(Y)|X = x] = \begin{cases} \sum_y h(y)p_{Y|X}(y|x), & \text{for } Y \text{ discrete,} \\ \int_{-\infty}^{\infty} h(y)f_{Y|X}(y|x)\,dy, & \text{for } Y \text{ continuous.} \end{cases}$$

Let $\mu(x)$ denote $E(Y|X = x)$. This is known as the regression function of Y on x. The regression function is often used to predict Y given a value $X = x$.

Definition: The conditional expectation of Y given X is the random variable $E(Y|X)$ which is equal to $E(Y|X = x)$ when $X = x$. Thus, $E(Y|X)$ is a rv that is a function of the rv X.

Example 12 again Obtain the conditional expectation of X given $Y = y$.

We earlier found the conditional pmf of X given $Y = 0$ is

$$p_{X|Y}(0|0) = \frac{1/16}{1/16} = 1, \qquad x = 0.$$

Thus, $E[X|Y = 0] = 1$.

The conditional pmf of X given $Y = 1$ is

$$p_{X|Y}(x|1) = \begin{cases} \frac{3/16}{4/16} = \frac{3}{4}, & x = 0 \\ \frac{1/16}{4/16} = \frac{1}{4}, & x = 1 \end{cases}$$

Thus, $E[X|Y = 1] = (3/4)(0) + (1/4)(1) = 1/4$.

The conditional pmf of X given $Y = 2$ is

$$p_{X|Y}(x|2) = \begin{cases} \frac{3/16}{6/16} = \frac{1}{2}, & x = 0 \\ \frac{2/16}{6/16} = \frac{1}{3}, & x = 1 \\ \frac{1/16}{6/16} = \frac{1}{6} & x = 2. \end{cases}$$

Thus, $E[X|Y = 2] = (0)(1/2) + (1)(1/3) + (2)(1/6) = 2/3$.

We can likewise compute the remaining conditional expectations:

$$E[X|Y=3] = (0)(1/4) + (1)(1/4) + (2)(1/4) + (3)(1/4) = 3/2$$

$$E[X|Y=4] = 4$$

We now consider the distribution of $W = E[X|Y]$. Recall from Chapter 2 that the marginal pmf of Y is

y	0	1	2	3	4
$p_Y(y)$	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$
$E[X\|Y=y]$	1	$1/4$	$2/3$	$3/2$	4

Thus, W is a discrete rv with pmf

w	$1/4$	$2/3$	1	$3/2$	4
$p_W(w)$	$4/16$	$6/16$	$1/16$	$4/16$	$1/16$

<u>Example 19 again</u> Let X and Y have joint pdf

$$f(x,y) = \begin{cases} 3(x+y), & 0 < x < 1, 0 < y < 1, \\ & 0 < x+y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Earlier we found that the marginal pdf of X was
$f_X(x) = \frac{3}{2}(1-x^2),\ 0 < x < 1.$

The conditional pdf of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{2(x+y)}{1-x^2},\ 0 < y < 1-x.$$

The regression function of Y on x is

$$E(Y|x) = \int_0^{1-x} y \frac{2(x+y)}{1-x^2} dy = \frac{2-x-x^2}{3(1+x)}$$

Example 28 again   Let X and Y have joint pdf

$$f(x, y) = \begin{cases} 2, & 0 < x < 1, 0 < y < 1, \\ & 0 < x + y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

The marginal pdf of X is

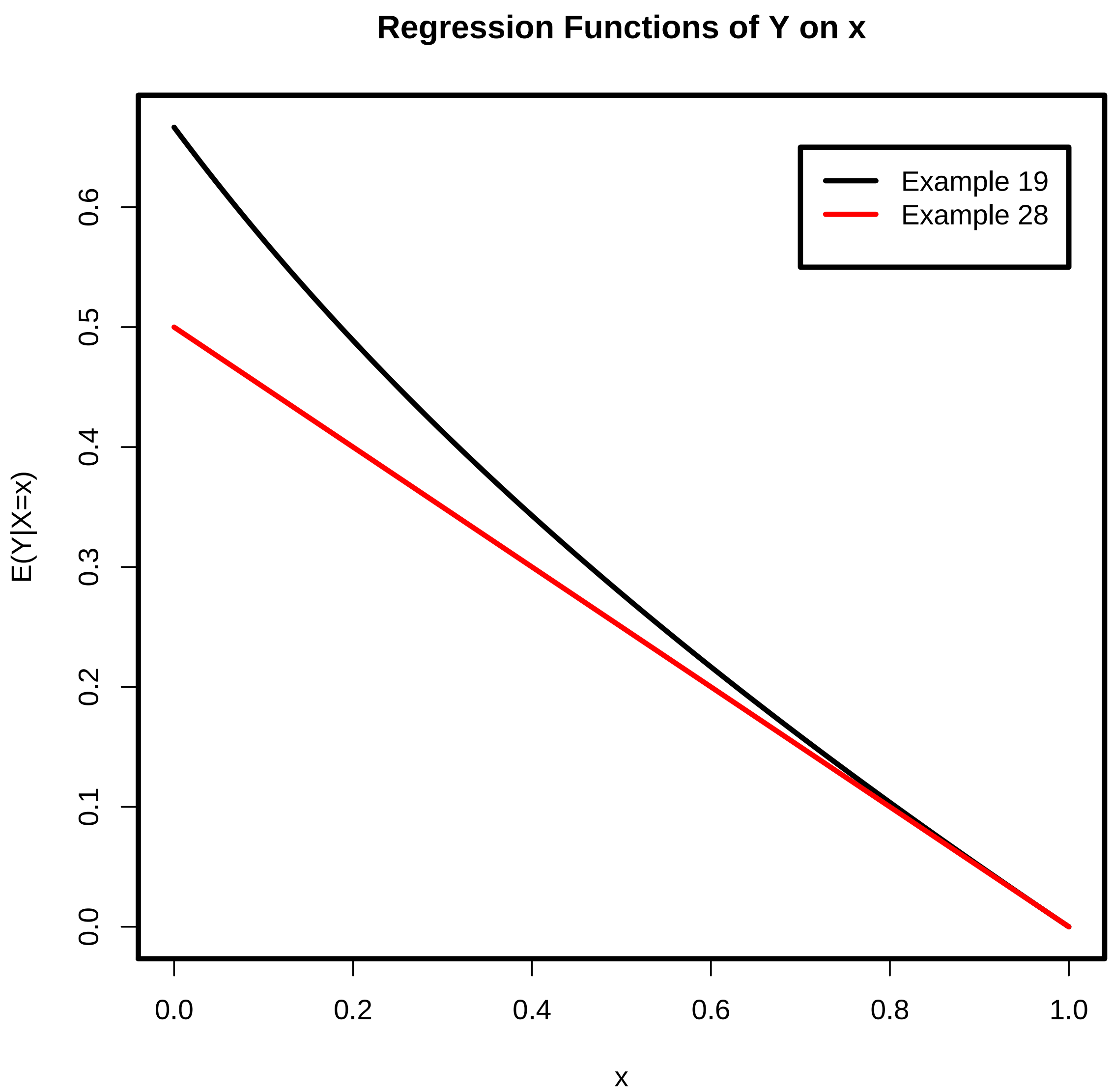
$$f_X(x) = \int_0^{1-x} 2dy = 2(1-x),\ 0 < x < 1.$$

The conditional pdf of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{2}{2(1-x)} = \frac{1}{1-x},\ 0 < y < 1 - x.$$

The regression function of Y on x is

$$E(Y|x) = \int_0^{1-x} y \, \frac{1}{1-x} dy = \frac{1-x}{2}$$


Regression Functions of Y on x
Example 19
Example 28
E(Y|X=x)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.0
0.2
0.4
0.6
0.8
1.0
x

6.1 Properties of Conditional Expectation and Conditional Variance

The conditional variance of Y given $X = x$, $\mathrm{Var}(Y|X = x)$, is defined to be

$$\mathrm{Var}(Y|X = x) = E[(Y - \mu(x))^2|X = x].$$

Theorem A: $E[E(Y|X)] = E(Y)$.

Theorem B: $\mathrm{Var}(Y) = \mathrm{Var}[E(Y|X)] + E[\mathrm{Var}(Y|X)]$

Example: Suppose that a particle counter is imperfect and independently detects each incoming particle with probability θ. Suppose that the distribution of number N of incoming particles is Poisson (λ). Then the conditional distribution of the number (X) of counted particles given $N = n$ is binomial (n, θ). It is an interesting exercise to show that the unconditional distribution of the counted number of particles X is Poisson $(\lambda\theta)$.

Example: We continue the particle example. We assumed that the conditional distribution of $X|N = n$ is binomial (n, θ) and that N is Poisson (λ). Find $E(X)$ and $\mathrm{Var}(X)$.

The conditional mean and variance of X given N are

$$E(X|N) = N\theta \qquad \text{and} \qquad \mathrm{Var}(X|N) = N\theta(1-\theta).$$

Then

$$E(X) = E[E(X|N)] = E[N\theta] = \lambda\theta$$

and

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{Var}[E(X|N)] + E[\mathrm{Var}(X|N)] \\ &= \mathrm{Var}[N\theta] + E[N\theta(1-\theta)] \\ &= \lambda\theta^2 + \lambda\theta(1-\theta) = \lambda\theta\end{aligned}$$

6.2 Inequalities

There are several useful probability inequalities that give bounds on probabilities using certain expectations. A basic inequality is Markov's inequality.

Markov's inequality: Let X be a random variable such that $P(X \geq 0) = 1$. Then for every positive number a,

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Proof: We'll assume X is a continuous rv with pdf f.

$$
\begin{aligned}
E(X) &= \int_0^\infty x f(x)\,dx \\
&= \int_0^a x f(x)\,dx + \int_a^\infty x f(x)\,dx \\
&\geq \int_0^a x f(x)\,dx + \int_a^\infty a f(x)\,dx \\
&= \int_0^a x f(x)\,dx + a \int_a^\infty f(x)\,dx \\
&= \int_0^a x f(x)\,dx + aP(X \geq a) \quad \geq \quad aP(X \geq a).
\end{aligned}
$$

Remark: The usefulness of Markov's inequality is that it allows us to say something about the whole distribution of X when all we know is the first moment.

6.3 Chebyshev's Inequality

Chebyshev's inequality: Suppose X is a random variable with finite variance σ_X^2, and let $\mu_X = E(X)$. Then for each $a > 0$

$$P(|X - \mu_X| \geq a) \leq \frac{\sigma_X^2}{a^2}.$$

Proof: We may write

$$P(|X - \mu_X| \geq a) = P\left[(X - \mu_X)^2 \geq a^2\right].$$

The result follows upon applying Markov's inequality to the rv $Y = (X - \mu_X)^2$ and using the fact that

$$E(Y) = E\left[(X - \mu_X)^2\right] = \sigma_X^2.$$

With Chebyshev's inequality, we may make at least crude determinations about a distribution when all we know are the first two moments of the distribution.

For example, take $a = 1.5$. Chebyshev's inequality tells us that

$$P(|X - \mu| \geq 1.5\sigma) \leq \frac{1}{1.5^2} = \frac{4}{9} < 0.45,$$

or

$$P(|X - \mu| < 1.5\sigma) > 0.55.$$

So, for **any** distribution with a finite variance, at least 55% of the distribution must lie within 1.5 standard deviations of the mean.

