11.2 Extending the Definitions to Multiple RVs

Definition 11.27. Joint pmf:

$$p_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = P\left[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\right].$$

Joint cdf:

$$F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = P\left[X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n\right].$$

11.28. Marginal pmf:

$$p_X(x) = \sum_y \sum_z p_{X,Y,Z}(x,y,z)$$

Example 11.29. Consider three random variables X, Y, and Z whose joint pmf is given by

$$p_{X,Y,Z}\left(x,y,z\right) = \begin{cases} 1/7, & (x,y,z) \in \left\{(0,1,0), (1,1,1)\right\}, \\ 2/7, & (x,y,z) = (0,0,1), \\ 3/7, & (x,y,z) = (0,1,1), \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$\begin{aligned} p_X\left(0\right) &\equiv P\left[X = 0\right] = \\ p_X\left(1\right) &\equiv P\left[X = 1\right] = \end{aligned}$$

Therefore,

$$p_X\left(x\right) = \begin{cases} , & x = 0, \\ , & x = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Definition 11.30. ***Identically distributed*** random variables: The following statements are equivalent.

(a) Random variables $X_1, X_2, \ldots$ are ***identically distributed***

(b) For every B, $P\left[X_j \in B\right]$ does not depend on j.

(c) $p_{X_i}(c) = p_{X_j}(c)$ for all c, i, j.

(d) $F_{X_i}(c) = F_{X_j}(c)$ for all c, i, j.

Definition 11.31. ***Independence*** among finite number of random variables: The following statements are equivalent.

(a) $X_1, X_2, \ldots, X_n$ are ***independent***

(b) $[X_1 \in B_1], [X_2 \in B_2], \ldots, [X_n \in B_n]$ are independent, for all $B_1, B_2, \ldots, B_n$.

(c) $P[X_i \in B_i, \forall i] = \prod_{i=1}^{n} P[X_i \in B_i]$, for all $B_1, B_2, \ldots, B_n$.

(d) $p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p_{X_i}(x_i)$ for all $x_1, x_2, \ldots, x_n$.

(e) $F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$ for all $x_1, x_2, \ldots, x_n$.

Example 11.32. Toss a coin n times. For the ith toss, let

$$X_i = \begin{cases} 1, & \text{if H happens on the } i\text{th toss,} \\ 0, & \text{if T happens on the } i\text{th toss.} \end{cases}$$

We then have a collection of i.i.d. random variables $X_1, X_2, X_3, \ldots, X_n$.

11.33. Fact: For i.i.d. $X_i \sim$ Bernoulli(p), $Y = X_1 + X_2 + \cdots + X_n$ is $\mathcal{B}(n, p)$.

To see this, consider n (independent) Bernoulli trials (as in Example 11.32). Let

$$X_i = \begin{cases} 1, & \text{if success happens on the } i\text{th trial,} \\ 0, & \text{if failure happens on the } i\text{th trial.} \end{cases}$$

Then, Y is simply counting the number of successes in the n trials. From Definition 8.33 of Binomial RV, we conclude that Y is binomial.

Example 11.34. Roll a dice n times. Let N_i be the result of the ith roll. We then have another collection of i.i.d. random variables $N_1, N_2, N_3, \ldots, N_n$.

Example 11.35. Let X_1 be the result of tossing a biased coin. Set $X_2 = X_3 = \cdots = X_n = X_1$.

11.36. If $X_1, X_2, \dots, X_n$ are independent, then so is any subcollection of them.

Definition 11.37. A ***pairwise independent*** collection of random variables is a collection of random variables any two of which are independent.

(a) Any collection of (mutually) independent random variables is pairwise independent

(b) Some pairwise independent collections are not independent. See Example (11.38).

Example 11.38. Let suppose X, Y, and Z have the following joint probability distribution: $p_{X,Y,Z}(x,y,z) = \frac{1}{4}$ for $(x,y,z) \in \{(0,0,0),(0,1,1),(1,0,1),(1,1,0)\}$. This, for example, can be constructed by starting with independent X and Y that are Bernoulli-$\frac{1}{2}$. Then set $Z = X \oplus Y = X + Y \mod 2$.

(a) X, Y, Z are pairwise independent.

(b) X, Y, Z are not independent.

11.3 Expectation of Function of Discrete Random Variables

11.39. Recall that the expected value of "any" function g of a discrete random variable X can be calculated from

$$\mathbb{E}[g(X)] = \sum_x g(x) p_X(x).$$

Similarly[53], the expected value of "any" function g of two discrete random variables X and Y can be calculated from

$$\mathbb{E}[g(X,Y)] = \sum_x \sum_y g(x,y) p_{X,Y}(x,y).$$

$\sum_{(x,y)}$

[53] Again, these are called the **law/rule of the lazy statistician** (LOTUS) [22, Thm 3.6 p 48],[9, p. 149] because it is so much easier to use the above formula than to first find the pmf of $g(X)$ or $g(X,Y)$. It is also called **substitution rule** [21, p 271].

	Discrete
$P[X \in B]$	$\sum_{x \in B} p_X(x)$
$P[(X,Y) \in R]$	$\sum_{(x,y):(x,y) \in R} p_{X,Y}(x,y)$
Joint to Marginal: (Law of Total Prob.)	$p_X(x) = \sum_y p_{X,Y}(x,y)$ $p_Y(y) = \sum_x p_{X,Y}(x,y)$
$P[X > Y]$	$\sum_x \sum_{y:\, y<x} p_{X,Y}(x,y)$ $= \sum_y \sum_{x:\, x>y} p_{X,Y}(x,y)$
$P[X = Y]$	$\sum_x p_{X,Y}(x,x)$
$X \perp\!\!\!\perp Y$	$p_{X,Y}(x,y) = p_X(x)p_Y(y)$
Conditional	$p_{X\|Y}(x\|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$
$\mathbb{E}[g(X,Y)]$	$\sum_x \sum_y g(x,y) p_{X,Y}(x,y)$

Table 8: Joint pmf: A Summary

11.40. $\mathbb{E}[\cdot]$ is a **linear** operator: $\mathbb{E}[aX + bY] = a\mathbb{E}X + b\mathbb{E}Y$.

$$\mathbb{E}[3X + 5Y] = 3\mathbb{E}X + 5\mathbb{E}Y$$

(a) Homogeneous: $\mathbb{E}[cX] = c\mathbb{E}X$

(b) Additive: $\mathbb{E}[X + Y] = \mathbb{E}X + \mathbb{E}Y$

(c) Extension: $\mathbb{E}\left[\sum_{i=1}^n c_i g_i(X_i)\right] = \sum_{i=1}^n c_i \mathbb{E}[g_i(X_i)]$.

$$\mathbb{E}[3X^2 + 8\sqrt{Y} + 5Z] = \mathbb{E}[3X^2] + \mathbb{E}[8\sqrt{Y}] + \mathbb{E}[5Z]$$
$$= 3\mathbb{E}[X^2] + 8\mathbb{E}[\sqrt{Y}] + 5\mathbb{E}[Z]$$

Example 11.41. Recall from 11.33 that when i.i.d. $X_i \sim$ Bernoulli(p), $Y = X_1 + X_2 + \cdots + X_n$ is $\mathcal{B}(n,p)$. Also, from Example 9.4, we have $\mathbb{E}X_i = p$. Hence,

$$\mathbb{E}Y = \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n p = np.$$

Therefore, the expectation of a binomial random variable with parameters n and p is np.

Example 11.42. A binary communication link has bit-error probability p. What is the expected number of bit errors in a transmission of n bits?

Theorem 11.43 (Expectation and Independence)**.** Two random variables X and Y are independent if and only if

$$\mathbb{E}\left[h(X)g(Y)\right] = \mathbb{E}\left[h(X)\right]\mathbb{E}\left[g(Y)\right]$$

for "all" functions h and g.

Ex. If $X \perp\!\!\!\perp Y$, then

$\mathbb{E}\left[(X^3-3)^2 \sin Y\right] = \mathbb{E}\left[(X^3-3)^2\right] \times \mathbb{E}\left[\sin Y\right]$

- In other words, X and Y are independent if and only if for every pair of functions h and g, the expectation of the product $h(X)g(Y)$ is equal to the product of the individual expectations.

- One special case is that

$$X \perp\!\!\!\perp Y \quad \text{implies} \quad \mathbb{E}\left[XY\right] = \mathbb{E}X \times \mathbb{E}Y. \tag{33}$$

  However, independence means more than this property. In other words, having $\mathbb{E}\left[XY\right] = (\mathbb{E}X)(\mathbb{E}Y)$ does not necessarily imply $X \perp\!\!\!\perp Y$. See Example 11.54.

11.44. Let's combined what we have just learned about independence into the definition/equivalent statements that we already have in 11.21.

The following statements are equivalent:

(a) Random variables X and Y are ***independent***.

(b) $[X \in B] \perp\!\!\!\perp [Y \in C]$ for all B, C. ← event-based defn.

(c) $P\left[X \in B, Y \in C\right] = P\left[X \in B\right] \times P\left[Y \in C\right]$ for all B, C.

(d) $p_{X,Y}(x, y) = p_X(x) \times p_Y(y)$ for all x, y.

(e) $F_{X,Y}(x, y) = F_X(x) \times F_Y(y)$ for all x, y.

(f) $\mathbb{E}\left[h(X)g(Y)\right] = \mathbb{E}\left[h(X)\right]\mathbb{E}\left[g(Y)\right]$ for all $h(\cdot), g(\cdot)$

(g) $I(X;Y) = 0$

mutual information (ECS 452)

$\text{Var}[Z] \equiv \mathbb{E}[(Z-\mathbb{E}Z)^2] = \mathbb{E}[Z^2] - (\mathbb{E}Z)^2 = \mathbb{E}[X^2Y^2] - (\mathbb{E}[XY])^2 = 9 - 1 = 8$

$\mathbb{E}[XY] = \mathbb{E}X\,\mathbb{E}Y = 1 \times 1 = 1$ $\quad$ $\mathbb{E}[X^2Y^2] = \mathbb{E}[X^2]\,\mathbb{E}[Y^2] = (\underbrace{\text{Var}[X]}_{2} + \underbrace{(\mathbb{E}X)^2}_{+1})(3) = 9$

Exercise 11.45 (F2011). Suppose X and Y are i.i.d. with $\mathbb{E}X = \mathbb{E}Y = 1$ and $\text{Var}\,X = \text{Var}\,Y = 2$. Find $\text{Var}[\underbrace{XY}_{Z}]$.

11.46. To quantify the amount of ***dependence*** between two random variables, we may calculate their ***mutual information***. This quantity is crucial in the study of digital communications and information theory. However, in introductory probability class (and introductory communication class), it is traditionally omitted.

11.4 Linear Dependence

Definition 11.47. Given two random variables X and Y, we may calculate the following quantities:

(a) **Correlation**: $\mathbb{E}[XY]$.

(b) **Covariance**: $\text{Cov}[X,Y] = \mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)]$.

(c) **Correlation coefficient**: $\rho_{X,Y} = \frac{\text{Cov}[X,Y]}{\sigma_X\sigma_Y}$

Exercise 11.48 (F2011). Continue from Exercise 11.7.

(a) Find $\mathbb{E}[XY]$.

(b) Check that $\text{Cov}[X,Y] = -\frac{1}{25}$.

Recall, if $X \perp\!\!\!\perp Y$, then
$\mathbb{E}[h(X)g(Y)] = \mathbb{E}[h(X)]\,\mathbb{E}[g(Y)]$
A special of ↗ is
$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$

11.49. $\text{Cov}[X,Y] = \mathbb{E}[(X-\overset{m_X}{\mathbb{E}X})(Y-\overset{m_Y}{\mathbb{E}Y})] = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y$ $= 0$ when $X \perp\!\!\!\perp Y$

$$= \mathbb{E}[XY - m_XY - m_YX + m_Xm_Y]$$
$$= \mathbb{E}[XY] - m_X\underbrace{\mathbb{E}Y}_{m_Y} - \cancel{m_Y\underbrace{\mathbb{E}X}_{m_X}} + \cancel{m_Xm_Y}$$

Compare this with
$\text{Var}[X] = \mathbb{E}[(X-\mathbb{E}X)^2]$
$= \mathbb{E}[X^2] - (\mathbb{E}X)^2$

- Note that $\text{Var}\,X = \text{Cov}[X,X]$.

11.50. $\text{Var}[\underbrace{X+Y}_{Z}] = \text{Var}\,X + \text{Var}\,Y + 2\text{Cov}[X,Y]$

$Z = X+Y$
$\mathbb{E}Z = \mathbb{E}X + \mathbb{E}Y$

$$\equiv \mathbb{E}[(Z-\mathbb{E}Z)^2] = \mathbb{E}[(X+Y-(\mathbb{E}X+\mathbb{E}Y))^2]$$
$$= \mathbb{E}[(\underbrace{(X-\mathbb{E}X)}_{A} + \underbrace{(Y-\mathbb{E}Y)}_{B})^2] = \mathbb{E}[(A+B)^2] = \mathbb{E}[A^2+B^2+2AB]$$
$$= \mathbb{E}[A^2] + \mathbb{E}[B^2] + 2\,\mathbb{E}[AB] = \mathbb{E}[(X-\mathbb{E}X)^2] + \mathbb{E}[(Y-\mathbb{E}Y)^2]$$
$$+ 2\,\mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)]$$

If X and Y are uncorrelated (or independent), $\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$

Definition 11.51. X and Y are said to be ***uncorrelated*** if and only if $\text{Cov}[X,Y] = 0$.

11.52. The following statements are equivalent:

(a) X and Y are ***uncorrelated***.

(b) $\text{Cov}[X,Y] = 0$.

(c) $\mathbb{E}[XY] = \mathbb{E}X\mathbb{E}Y$.

(d)

11.53. Independence implies uncorrelatedness; that is if $X \perp\!\!\!\perp Y$, then $\text{Cov}[X,Y] = 0$.

The converse is not true. Uncorrelatedness does not imply independence. See Example 11.54.

Example 11.54. Let X be uniform on $\{\pm 1, \pm 2\}$ and $Y = |X|$.

$y = \lvert x \rvert$	x	$p_X(x)$
2	-2	1/4
1	-1	1/4
1	1	1/4
2	2	1/4

$$p_Y(y) = \begin{cases} 1/2, & y = 1, 2, \\ 0, & \text{otherwise}. \end{cases}$$

$P[Y=1] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

$P[Y=2] = 1 - \frac{1}{2} = \frac{1}{2}$

$$\mathbb{E}[X] = \sum_x x\, p_X(x) = \cancel{(-2)}\frac{1}{4} + \cancel{(-1)}\frac{1}{4} + \cancel{(1)}\frac{1}{4} + \cancel{(2)}\frac{1}{4} = 0$$

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \mathbb{E}[|X|] = \sum_x |x|\, p_X(x) = (2)\times\frac{1}{4} + (1)\times\frac{1}{4} + (1)\times\frac{1}{4} + (2)\times\frac{1}{4} = \frac{3}{2} = 1.5$$

$$\mathbb{E}[Y] = \sum_y y\, p_Y(y) = (1)\frac{1}{2} + (2)\frac{1}{2} = \frac{3}{2} = 1.5$$

$$\mathbb{E}[XY] = \sum_x \sum_y (xy)\, p_{X,Y}(x,y) = \cancel{(-4)}\frac{1}{4} + \cancel{(-1)}\frac{1}{4} + \cancel{(1)}\frac{1}{4} + \cancel{(4)}\frac{1}{4} = 0$$

$$\text{Cov}[X,Y] = \mathbb{E}[XY] - \mathbb{E}X\,\mathbb{E}Y = 0 - (0)(1.5) = 0$$

↳ uncorrelated

$$\begin{array}{c|cc} x \backslash y & 1 & 2 \\ \hline -2 & 0 & 1/4 \\ -1 & 1/4 & 0 \\ 1 & 1/4 & 0 \\ 2 & 0 & 1/4 \end{array}$$

$$p_{X,Y}(-2,2) = P[\underbrace{X=-2}_{A}, \underbrace{Y=2}_{B}] = \overbrace{P[X=-2]}^{P(A)\ =\ 1/4}\ \underbrace{P[Y=2 \mid X=-2]}_{P(B|A)\ =\ 1}$$

P(A ∩ B)

$p_{X,Y}(1,1) = \frac{1}{4}$

$p_X(1) = \frac{1}{4}$

$p_Y(1) = \frac{1}{2}$

11.55. The variance of the sum of uncorrelated (or independent) random variables is the sum of their variances.

$p_{X,Y}(1,1) \neq p_X(1)\, p_Y(1)$

$\Rightarrow$ not indp.

Exercise 11.56. Suppose two fair dice are tossed. Denote by the random variable V_1 the number appearing on the first dice and by the random variable V_2 the number appearing on the second dice. Let $X = V_1 + V_2$ and $Y = V_1 - V_2$.

(a) Show that X and Y are not independent.

(b) Show that $\mathbb{E}[XY] = \mathbb{E}X\mathbb{E}Y$.

11.57. $\text{Cov}[aX + b, cY + d] = ac\text{Cov}[X, Y]$

$$\begin{aligned}
\text{Cov}[aX + b, cY + d] &= \mathbb{E}[((aX + b) - \mathbb{E}[aX + b])((cY + d) - \mathbb{E}[cY + d])] \\
&= \mathbb{E}[((aX + b) - (a\mathbb{E}X + b))((cY + d) - (c\mathbb{E}Y + d))] \\
&= \mathbb{E}[(aX - a\mathbb{E}X)(cY - c\mathbb{E}Y)] \\
&= ac\mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] \\
&= ac\text{Cov}[X, Y].
\end{aligned}$$

Definition 11.58. ***Correlation coefficient***:

$$\begin{aligned}
\rho_{X,Y} &= \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y} \\
&= \mathbb{E}\left[\left(\frac{X - \mathbb{E}X}{\sigma_X}\right)\left(\frac{Y - \mathbb{E}Y}{\sigma_Y}\right)\right] = \frac{\mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y}{\sigma_X \sigma_Y}.
\end{aligned}$$

normalized/standardized version of X.

- $\rho_{X,Y}$ is dimensionless
- $\rho_{X,X} = 1$
- $\rho_{X,Y} = 0$ if and only if X and Y are uncorrelated.
- ***Cauchy-Schwartz Inequality***[54]:

$$|\rho_{X,Y}| \leq 1.$$

In other words, $\rho_{XY} \in [-1, 1]$.

[54]Cauchy-Schwartz inequality shows up in many areas of Mathematics. A general form of this inequality can be stated in any inner product space:

$$|\langle a, b \rangle|^2 \leq \langle a, a \rangle \langle b, b \rangle.$$

Here, the inner product is defined by $\langle X, Y \rangle = \mathbb{E}[XY]$. The Cauchy-Schwartz inequality then gives

$$|\mathbb{E}[XY]|^2 \leq \mathbb{E}[X^2]\,\mathbb{E}[Y^2].$$

11.59. Linear Dependence and Cauchy-Schwartz Inequality

(a) If $Y = aX + b$, then $\rho_{X,Y} = \text{sign}(a) = \begin{cases} 1, & a > 0 \\ -1, & a < 0. \end{cases}$

- To be rigorous, we should also require that $\sigma_X > 0$ and $a \neq 0$.

(b) When $\sigma_Y, \sigma_X > 0$, equality occurs in the Cauchy-Schwartz inequality if and only if the following conditions holds

$$\begin{aligned} &\equiv \exists a \neq 0 \text{ such that } (X - \mathbb{E}X) = a(Y - \mathbb{E}Y) \\ &\equiv \exists a \neq 0 \text{ and } b \in \mathbb{R} \text{ such that } X = aY + b \\ &\equiv \exists c \neq 0 \text{ and } d \in \mathbb{R} \text{ such that } Y = cX + d \\ &\equiv |\rho_{XY}| = 1 \end{aligned}$$

In which case, $|a| = \frac{\sigma_X}{\sigma_Y}$ and $\rho_{XY} = \frac{a}{|a|} = \operatorname{sgn} a$. Hence, ρ_{XY} is used to quantify **linear dependence** between X and Y. The closer $|\rho_{XY}|$ to 1, the higher degree of linear dependence between X and Y.

Example 11.60. [21, Section 5.2.3] Consider an important fact that *investment experience* supports: spreading investments over a variety of funds (diversification) diminishes risk. To illustrate, imagine that the random variable X is the return on every invested dollar in a local fund, and random variable Y is the return on every invested dollar in a foreign fund. Assume that random variables X and Y are i.i.d. with expected value 0.15 and standard deviation 0.12.

If you invest all of your money, say c, in either the local or the foreign fund, your return R would be cX or cY.

- The expected return is $\mathbb{E}R = c\mathbb{E}X = c\mathbb{E}Y = 0.15c$.
- The standard deviation is $c\sigma_X = c\sigma_Y = 0.12c$

Now imagine that your money is equally distributed over the two funds. Then, the return R is $\frac{1}{2}cX + \frac{1}{2}cY$. The expected return is $\mathbb{E}R = \frac{1}{2}c\mathbb{E}X + \frac{1}{2}c\mathbb{E}Y = 0.15c$. Hence, the expected return remains at 15%. However,

$$\operatorname{Var} R = \operatorname{Var}\left[\frac{c}{2}(X+Y)\right] = \frac{c^2}{4}\operatorname{Var} X + \frac{c^2}{4}\operatorname{Var} Y = \frac{c^2}{2} \times 0.12^2.$$

So, the standard deviation is $\frac{0.12}{\sqrt{2}}c \approx 0.0849c$.

In comparison with the distributions of X and Y, the pmf of $\frac{1}{2}(X+Y)$ is concentrated more around the expected value. The centralization of the distribution as random variables are averaged together is a manifestation of the central limit theorem.

11.61. [21, Section 5.2.3] Example 11.60 is based on the assumption that return rates X and Y are independent from each other. In the world of investment, however, risks are more commonly reduced by combining negatively correlated funds (two funds are negatively correlated when one tends to go up as the other falls).

This becomes clear when one considers the following hypothetical situation. Suppose that two stock market outcomes ω_1 and ω_2 are possible, and that each outcome will occur with a probability of $\frac{1}{2}$ Assume that domestic and foreign fund returns X and Y are determined by $X(\omega_1) = Y(\omega_2) = 0.25$ and $X(\omega_2) = Y(\omega_1) = -0.10$. Each of the two funds then has an expected return of 7.5%, with equal probability for actual returns of 25% and -10%. The random variable $Z = \frac{1}{2}(X+Y)$ satisfies $Z(\omega_1) = Z(\omega_2) = 0.075$. In other words, Z is equal to 0.075 with certainty. This means that an investment that is equally divided between the domestic and foreign funds has a guaranteed return of 7.5%.