Sirindhorn International Institute of Technology
Thammasat University
School of Information, Computer and Communication Technology

ECS315 2012/1 Part III.4 Dr.Prapun

10.3 Function of Discrete Random Variables

10.37. Recall that for discrete random variable X, the pmf of a derived random variable $Y = g(X)$ is given by

$$p_Y(y) = \sum_{x:g(x)=y} p_X(x).$$

Similarly, for discrete random variables X and Y, the pmf of a derived random variable $Z = g(X, Y)$ is given by

$$P[g(X,Y)=z] = P[Z=z] = p_Z(z) = \sum_{(x,y):g(x,y)=z} p_{X,Y}(x, y).$$

$$= \sum_x \sum_{y:g(x,y)=z} p_{X,Y}(x,y) = \sum_y \sum_{x:g(x,y)=z} p_{X,Y}(x,y)$$

Example 10.38. Suppose the joint pmf of X and Y is given by

$$p_{X,Y}(x, y) = \begin{cases} 1/15, & x = 0, y = 0, \\ 2/15, & x = 1, y = 0, \\ 4/15, & x = 0, y = 1, \\ 8/15, & x = 1, y = 1, \\ 0, & \text{otherwise.} \end{cases}$$

$$p_{X,Y}: \quad \begin{array}{c|cc} x \backslash y & 0 & 1 \\ \hline 0 & 1 & 4 \\ 1 & 2 & 8 \end{array} \times \frac{1}{15}$$

Let $Z = X + Y$. Find the pmf of Z.

$$x+y: \quad \begin{array}{c|cc} x \backslash y & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 2 \end{array}$$

$$p_Z(z) = \begin{cases} 1/15, & z = 0, \\ 2/5, & z = 1, \\ 8/15, & z = 2, \\ 0, & \text{otherwise.} \end{cases}$$

Exercise 10.39 (F2011)**.** Continue from Exercise 10.15. Let $Z = X + Y$.

(a) Find the pmf of Z.

(b) Find $\mathbb{E}Z$.

10.40. In general, when $Z = X + Y$,

$$\begin{aligned} p_Z(z) &= \sum_{(x,y):x+y=z} p_{X,Y}(x,y) \\ &= \sum_y p_{X,Y}(z-y,y) = \sum_x p_{X,Y}(x,z-x). \end{aligned}$$

Furthermore, if X and Y are independent,

$$p_Z(z) = \sum_{(x,y):x+y=z} p_X(x)\, p_Y(y) \tag{20}$$

$$= \sum_y p_X(z-y)\, p_Y(y) = \sum_x p_X(x)\, p_Y(z-x). \tag{21}$$

Example 10.41. Suppose $\Lambda_1 \sim \mathcal{P}(\lambda_1)$ and $\Lambda_2 \sim \mathcal{P}(\lambda_2)$ are independent. Let $\Lambda = \Lambda_1 + \Lambda_2$. Use (21) to show[33] that $\Lambda \sim \mathcal{P}(\lambda_1 + \lambda_2)$.

First, note that $p_\Lambda(\ell)$ would be positive only on nonnegative integers because a sum of nonnegative integers (Λ_1 and Λ_2) is still a nonnegative integer. So, the support of Λ is the same as the support for Λ_1 and Λ_2. Now, we know, from (21), that

$$P[\Lambda = \ell] = P[\Lambda_1 + \Lambda_2 = \ell] = \sum_i P[\Lambda_1 = i]\, P[\Lambda_2 = \ell - i]$$

Of course, we are interested in ℓ that is a nonnegative integer. The summation runs over $i = 0, 1, 2, \ldots$. Other values of i would make $P[\Lambda_1 = i] = 0$. Note also that if $i > \ell$, then $\ell - i < 0$ and $P[\Lambda_2 = \ell - i] = 0$. Hence, we conclude that the index i can only

[33]Remark: You may feel that simplifying the sum in this example (and in Exercise 10.42 is difficult and tedious, in Section 14, we will introduce another technique which will make the answer obvious. The idea is to realize that (21) is a convolution and hence we can use Fourier transform to work with a product in another domain.

be integers from 0 to k:

$$\begin{aligned}
P[\Lambda=\ell] &= \sum_{i=0}^{\ell} e^{-\lambda_1}\frac{\lambda_1^i}{i!}e^{-\lambda_2}\frac{\lambda_2^{\ell-i}}{(\ell-i)!}\\
&= e^{-(\lambda_1+\lambda_2)}\frac{1}{\ell!}\sum_{i=0}^{\ell}\frac{\ell!}{i!\,(\ell-i)!}\lambda_1^i\lambda_2^{\ell-i}\\
&= e^{-(\lambda_1+\lambda_2)}\frac{1}{\ell!}\sum_{i=0}^{\ell}\lambda_1^i\lambda_2^{\ell-i} = e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^\ell}{\ell!},
\end{aligned}$$

where the last equality is from the binomial theorem. Hence, the sum of two independent Poisson random variables is still Poisson!

$$p_\Lambda(\ell) = \begin{cases} e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^\ell}{\ell!}, & \ell\in\{0,1,2,\dots\}\\ 0, & \text{otherwise}\end{cases}$$

check this. →

Exercise 10.42. Suppose $B_1 \sim \mathcal{B}(n_1,p)$ and $B_2 \sim \mathcal{B}(n_2,p)$ are independent. Let $B = B_1 + B_2$. Use (21) to show that $B \sim \mathcal{B}(n_1+n_2,p)$.

10.4 Expectation of Function of Discrete Random Variables

10.43. 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[34], the expected value of "any" function g of two discrete random variable X and Y can be calculated from

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

[34] Again, these are called the **law/rule of the lazy statistician** (LOTUS) [23, 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** [22, p 271].

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

Table 4: Joint pmf: A Summary

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

$$\mathbb{E}[aX+bY] = \sum_x \sum_y (ax+by) p_{X,Y}(x,y) = a \sum_x \sum_y x\, p_{X,Y}(x,y) + b \sum_y \sum_x y\, p_{X,Y}(x,y)$$

$$= a \underbrace{\sum_x x\, p_X(x)}_{\mathbb{E}X} + b \underbrace{\sum_y y\, p_Y(y)}_{\mathbb{E}Y}$$

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

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

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

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

$\mathcal{B}(1,p)$

Example 10.45. Recall from 10.34 that when i.i.d. $X_i \sim$ Bernoulli(p), $Y = X_1 + X_2 + \cdots X_n$ is $\mathcal{B}(n,p)$ (binomial). Also, from Example 9.4, we have $\mathbb{E}X_i = p$. Hence,

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

Compare this with $\sum_{y=0}^{n} y \binom{n}{y} p^y (1-p)^{n-y}$

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

Example 10.46. A binary communication link has bit-error probability p. What is the expected number of bit errors in a transmission of n bits?

np

Theorem 10.47 (Expectation and Independence)**.** Two random variables X and Y are independent if and only if

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

for all functions h and g.

$X \perp\!\!\!\perp Y$

$\mathbb{E}[(X^2-3)^4 \sin(Y)] = \mathbb{E}[(X^2-3)^4]\,\mathbb{E}[\sin Y]$

- 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

$$X \perp\!\!\!\perp Y \quad \text{implies} \quad \mathbb{E}[XY] = \mathbb{E}X \times \mathbb{E}Y. \tag{22}$$

However, independence means more than this property. In other words, having $\mathbb{E}[XY] = (\mathbb{E}X)(\mathbb{E}Y)$ does not necessarily imply $X \perp\!\!\!\perp Y$. See Example 10.59.

10.48. Let's combined what we have just learned about independence into the definition/equivalent statements that we already have in 10.29.

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.

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

$P[X \le x, Y \le y] =$

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

Exercise 10.49 (F2011)**.** Suppose X and Y are i.i.d. with $\mathbb{E}X = \mathbb{E}Y = 1$ and $\operatorname{Var} X = \operatorname{Var} Y = 2$. Find $\operatorname{Var}[XY]$. $= \mathbb{E}\left[(XY - \mathbb{E}[XY])^2\right]$

10.50. To quantify the amount of ***dependence*** between two random variables, we may calculate their ***mutual information***. $I(X;Y)$ 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.

(Affine)

10.5 Linear Dependence

Definition 10.51. Given two random variables X and Y, we may calculate the following quantities:

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

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

(c) **Correlation coefficient**: $\rho_{X,Y} = \frac{\operatorname{Cov}[X,Y]}{\sigma_X \sigma_Y}$ $= \mathbb{E}\left[\frac{X - \mathbb{E}X}{\sigma_X} \times \frac{Y - \mathbb{E}Y}{\sigma_Y}\right]$

Exercise 10.52 (F2011)**.** Continue from Exercise 10.15.

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

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

10.53. $\operatorname{Cov}[X,Y] = \mathbb{E}[(X - \underbrace{\mathbb{E}X}_{m_X})(Y - \underbrace{\mathbb{E}Y}_{m_Y})] = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y$

$\mathbb{E}[XY - m_X Y - m_Y X + m_X m_Y]$

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

10.54. $\operatorname{Var}[\overbrace{X+Y}^{Z}] = \operatorname{Var} X + \operatorname{Var} Y + 2\operatorname{Cov}[X,Y]$

$\mathbb{E}[(Z - \mathbb{E}Z)^2] = \mathbb{E}\left[\left((X+Y) - \mathbb{E}[X+Y]\right)^2\right] = \mathbb{E}\left[\left((X - \mathbb{E}X) + (Y - \mathbb{E}Y)\right)^2\right]$

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

Definition 10.55. X and Y are said to be ***uncorrelated*** if and only if $\text{Cov}[X,Y] = 0$.

10.56. 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) $\mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)] = 0$

10.57. 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 10.59.

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

Example 10.59. Let X be uniform on $\{\pm 1, \pm 2\}$ and $Y = |X|$.

$p_{X,Y}$

$x \backslash y$	1	2
-2	0	1/4
-1	1/4	0
1	1/4	0
2	0	1/4

$\underbrace{p_{X,Y}(-2,1)}_{0} \neq \underbrace{p_X(-2)}_{\frac{1}{4}} \underbrace{p_Y(1)}_{\frac{1}{2}}$ → not independent

$\mathbb{E}[XY] = \mathbb{E}[X|X|] = 0$, where $\mathbb{E}[X|X|] = \sum_x x|x| p_X(x)$

$\mathbb{E}[X] = \sum_x x\, p_X(x) = 0$

$\underbrace{\mathbb{E}[X]}_{=0}\mathbb{E}[Y] = 0$

$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ → uncorrelated

Exercise 10.60. 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 $\mathbb{E}[XY] = \mathbb{E}X\mathbb{E}Y$.

10.61. ***Cauchy-Schwartz Inequality***:

$$(\text{Cov}[X,Y])^2 \le \sigma_X^2 \sigma_Y^2$$

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

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

Definition 10.63. ***Correlation coefficient***:

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

- $\rho_{X,Y}$ is dimensionless
- $\rho_{X,X} = 1$
- $\rho_{X,Y} = 0$ if and only if X and Y are uncorrelated.

10.64. Linear Dependence and Cauchy-Schwartz Inequality

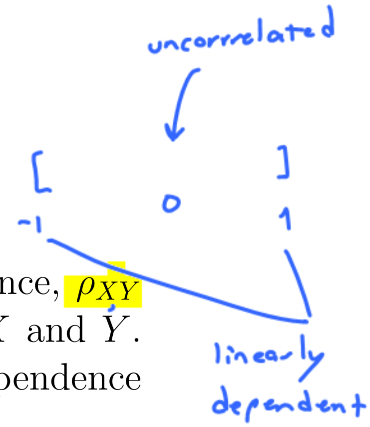
(a) If $Y = aX + b$, then $\rho_{X,Y} = \text{sign}(a) = \begin{cases} 1, & a > 0 \\ -1, & a < 0. \end{cases}$

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

(b) ***Cauchy-Schwartz Inequality***: $|\rho_{X,Y}| \le 1$.
In other words, $\rho_{XY} \in [-1, 1]$.

(c) When $\sigma_Y, \sigma_X > 0$, equality occurs if and only if the following conditions holds

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



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 10.65. [22, 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.

10.66. [22, Section 5.2.3] Example 10.65 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%.

Exercise 10.67. The input X and output Y of a system subject to random perturbations are described probabilistically by the following joint pmf matrix:

x \ y	2	4	5
1	0.02	0.10	0.08
3	0.08	0.32	0.40

(a) Evaluate the following quantities.

(i) $\mathbb{E}X$

(ii) $P[X = Y]$

(iii) $P[XY < 6]$

(iv) $\mathbb{E}[(X-3)(Y-2)]$

(v) $\mathbb{E}\left[X(Y^3 - 11Y^2 + 38Y)\right]$

(vi) $\operatorname{Cov}[X, Y]$

(vii) $\rho_{X,Y}$

(b) Calculate the following quantities using what you got from part (a).

(i) $\operatorname{Cov}[3X + 4, 6Y - 7]$

(ii) $\rho_{3X+4,6Y-7}$

(iii) $\operatorname{Cov}[X, 6X - 7]$

(iv) $\rho_{X,6X-7}$

Answers:

(a)

(i) $\mathbb{E}X = 2.6$

(ii) $P[X = Y] = 0$

(iii) $P[XY < 6] = 0.2$

(iv) $\mathbb{E}[(X - 3)(Y - 2)] = -0.88$

(v) $\mathbb{E}\left[X(Y^3 - 11Y^2 + 38Y)\right] = 104$

(vi) $\text{Cov}[X, Y] = 0.032$

(vii) $\rho_{X,Y} = 0.0447$

(b)

(i) Hence, $\text{Cov}[3X + 4, 6Y - 7] = 3 \times 6 \times \text{Cov}[X, Y] \approx 3 \times 6 \times 0.032 \approx \boxed{0.576}$.

(ii) Note that

$$\rho_{aX+b,cY+d} = \frac{\text{Cov}[aX + b, cY + d]}{\sigma_{aX+b}\sigma_{cY+d}}$$
$$= \frac{ac\text{Cov}[X, Y]}{|a|\sigma_X|c|\sigma_Y} = \frac{ac}{|ac|}\rho_{X,Y} = \text{sign}(ac) \times \rho_{X,Y}.$$

Hence, $\rho_{3X+4,6Y-7} = \text{sign}(3 \times 4)\rho_{X,Y} = \rho_{X,Y} = \boxed{0.0447}$.

(iii) $\text{Cov}[X, 6X - 7] = 1 \times 6 \times \text{Cov}[X, X] = 6 \times \text{Var}[X] \approx \boxed{3.84}$.

(iv) $\rho_{X,6X-7} = \text{sign}(1 \times 6) \times \rho_{X,X} = \boxed{1}$.