

# 1 Sums of independent r.v.& Symmetry

## 1.1 Convolution of two distributions

given two independent r.v.  $X$  and  $Y$ , the distribution of  $Z = X + Y$  is the convolution of the distributions of  $X$  and  $Y$ .

1. when  $X$  and  $Y$  are both discrete, the pmf of  $X + Y$  is given by

$$p_{X+Y}(n) = p_X * p_Y(n) = \sum_k p_{X(k)} p_{Y(n-k)} = \sum_k p_{X(n-k)} p_{Y(k)} \quad (1)$$

2. when  $X$  and  $Y$  are both continuous, the pdf of  $X + Y$  is given by

$$f_{X+Y}(z) = f_X * f_Y(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy \quad (2)$$

• *example: convolution of geometric random variables*

let  $X$  and  $Y$  be independent geometric random variables with the same success parameter  $p < 1$ , find the distribution of  $Z = X + Y$ .

We know  $p_X(k) = p_Y(k) = p(1-p)^{k-1}$   $k \geq 1$  r.v.  $Z = X + Y$  takes on values  $n = 2, 3, \dots$ . Via the convolution magic promised above, we have

$$\begin{aligned} P(X + Y = n) &= \sum_{k=-\infty}^{\infty} P(X = k) P(Y = n - k) \\ &= \sum_{k=1}^{n-1} p(X = k) P(Y = n - k) \\ &= \sum_{k=1}^{n-1} p(1-p)^{k-1} p(1-p)^{n-k-1} \\ &= \sum_{k=1}^{n-1} p^2 (1-p)^{n-2} \\ &= (n-1) p^2 (1-p)^{n-2} \end{aligned} \quad (3)$$

## 1.2 Negative binomial distribution

Coming off from the geometric distribution, we have the negative binomial distribution, which is the distribution of the number of trials needed to get  $r$  successes in a sequence of independent Bernoulli trials with success probability  $p$ . Its distribution, i.e. pmf, is given by

$$P(X = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k} \quad (n \geq k) \quad (4)$$

abbreviate this by  $X \sim \text{Negbin}(k, p)$ , where the geometric is a special case with  $k = 1$ .

## 1.3 Collection of normal distributed r.v.s

For  $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ ,  $X = \sum_i a_i X_i$ , we know

$$\begin{aligned} X &\sim \mathcal{N}(\mu, \sigma^2) \\ \text{where } \mu &= \sum_i a_i \mu_i, \sigma^2 = \sum_i a_i^2 \sigma_i^2 \end{aligned} \quad (5)$$

in other words, the sum of normal distributed r.v.s is also normal distributed.

## 1.4 Exchangeable r.v.s

a sequence of r.v.s  $X_1, X_2, X_3, \dots, X_n$  is **exchangeable** if the following condition holds: for any permutation  $(k_1, k_2, k_3)$  of  $(1, 2, \dots, n)$ , we have

$$(X_1, X_2, \dots, X_n) \stackrel{d}{=} (X_{k_1}, X_{k_2}, \dots, X_{k_n}) \quad (6)$$

### • How to check exchangeability

“it just works” method: check if the r.v. are identically distributed, i.e. if marginal pdf or pmf is the same.

Suppose  $X_1, X_2, \dots, X_n$  are discrete random variables with joint probability mass function  $p$ . Then these random variables are exchangeable if and only if  $p$  is a symmetric function.

Suppose  $X_1, X_2, \dots, X_n$  are jointly continuous random variables with joint density function  $f$ . Then these random variables are exchangeable if and only if  $f$  is a symmetric function.

If the expectation is conserved under permutations of our set of r.v.s.

**Importantly, if the r.v.s are independent and identically distributed, they are also exchangeable.**

remarks:

1. r.v. denoting outcomes of sampling without replacement  $X_1, X_2, \dots, X_n$  are exchangeable.
2. For any function  $g$  dependent on , the r.v.s  $g(X_1), g(X_2), \dots, g(X_n)$  are exchangeable.

## 1.5 Expectation and Variance of Multivariable r.v.

### 1.5.1 Expectation: linear

$$E[g_1(X_1) + g_2(X_2) + \dots + g_n(X_n)] = E[g_1(X_1)] + E[g_2(X_2)] + \dots + E[g_n(X_n)] \quad (7)$$

$$E[X_1 + X_2 + \dots + X_n] = E[X_1] + E[X_2] + \dots + E[X_n] \quad (8)$$

Expectation of a sum is always the sum of expectations.

### 1.5.2 Variance: sum of independent r.v., linear

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n) \quad (9)$$

### 1.5.3 The indicator method

- *example* We draw five cards from a deck of 52 without replacement. Let  $X$  denote the number of Aces among the chosen cards. Find the expected value of  $X$ .

Two ways to solve this:

1. Since order does not matter in our draw of 5, by argument of exchangeability, we can construct the following indicator:

$$I_i = \begin{cases} 1 & \text{if the } i\text{th card is an ace} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

Since  $X$  is the number of Aces among our 5 cards, we have

$$X = I_1 + I_2 + I_3 + I_4 + I_5 \quad (11)$$

Recall the linearity of expectation, we can rephrase the expected value as

$$E[X] = E[I_1] + E[I_2] + E[I_3] + E[I_4] + E[I_5] \quad (12)$$

Since r.v.  $I_i$  are exchangeable, we have

$$E[I_1] = E[I_2] = E[I_3] = E[I_4] = E[I_5] \quad (13)$$

Equation 12 becomes

$$5 * E[I_1] = 5 * P(I_1 = 1) = 5 * \frac{4}{52} = \frac{5}{13} \quad (14)$$

2. We can also label the Aces in the total deck as 1,2,3,4, and have our indicators  $j_1, j_2, j_3, j_4$  indicating if the  $i$ th Ace is in our draw or not. The number of Aces in our draw is then  $X = j_1 + j_2 + j_3 + j_4$ . By similar arguments of exchangeability, we have  $E[X] = 4E[j_1] = 4P(\text{one of the ace is among the 5 cards})$ . Notice that

$$\begin{aligned} P(\text{one of the ace is among the 5 cards}) &= \frac{\binom{1}{1}, \binom{51}{4}}{\binom{52}{5}} = \frac{5}{52} \\ \Rightarrow E[X] &= \frac{5}{13} \end{aligned} \quad (15)$$

#### 1.5.4 Expectation of multiple products

let  $X_1, X_2, X_3$  be independent r.v., when for all function  $g_1, g_2, g_3$

$$E\left[\prod_{i=1}^3 g_{i(X_i)}\right] = \prod_{i=1}^3 E[g_{i(X_i)}] \quad (16)$$

#### 1.6 Moment generating function with sums of r.v.

For independent r.v.  $X, Y$ , and mgf  $M_X(t), M_Y(t)$ ,

$$M_{X+Y}(t) = M_X(t)M_Y(t) \quad (17)$$

#### 1.7 Covariance and correlation

##### 1.7.1 Covariance

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y] \quad (18)$$

- $X$  &  $Y$  are
  - positively correlated if  $\text{Cov}(X, Y) > 0$
  - negatively correlated if  $\text{Cov}(X, Y) < 0$
  - uncorrelated if  $\text{Cov}(X, Y) = 0$

##### 1.7.2 Properties of Covariance

- $\text{COV}(X, Y) = \text{COV}(Y, X)$
- $\text{COV}(aX + b, Y) = a \text{COV}(X, Y)$
- for any r.v.  $X_i, Y_j$  and real numbers  $a_i, b_j$ :

$$\text{COV}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{COV}(X_i, Y_j) \quad (19)$$

- Practically,

$$\begin{aligned} \text{Cov}(Y_1 + Y_2, Z) &= \text{Cov}(Y_1, Z) + \text{Cov}(Y_2, Z) \\ \text{Cov}(X, X) &= \text{Var}(X) \end{aligned} \quad (20)$$

##### 1.7.3 Variance of sum of r.v.s

$$\text{Var} \left( \sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var} (X_i) + 2 \sum_{i \leq i < j \leq n} \text{Cov}(X_i, X_j) \quad (21)$$

For two r.v.s, this comes down to

$$\text{Var} (X + Y) = \text{Var} (x) + \text{Var} (Y) + 2 \text{Cov}(X, Y) \quad (22)$$

For three r.v.s, this is uglier...

$$\begin{aligned} & \text{Var} (X + Y + Z) \\ &= \text{Var} (X) + \text{Var} (Y) + \text{Var} (Z) + 2 \text{Cov}(X, Y) + 2 \text{Cov}(X, Z) + 2 \text{Cov}(Y, Z) \end{aligned} \quad (23)$$

You dont want to compute this for four or more...

#### 1.7.4 Correlation

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} \quad (24)$$

## 2 Tail bounds and limit theorems

### 2.1 Markov's inequality

For any non-negative r.v.  $X$  and any  $a > 0$ , we have

$$P(X \geq a) \leq \frac{E[X]}{a} \quad (25)$$

### 2.2 Chebyshev's inequality

For any r.v.  $X$  with finite mean and variance, and any  $k > 0$ , we have

$$P(|X - E[X]| \geq k) \leq \frac{\text{Var}(X)}{k^2} \quad (26)$$

normally used to find  $P(X \geq c + \mu) \leq \frac{\sigma^2}{c^2}$  and  $P(X \leq \mu - c) \leq \frac{\sigma^2}{c^2}$

### 2.3 generalized Law of large numbers

For a sequence of iid r.v.s  $X_1, X_2, \dots, X_n$  with finite mean  $E[X_i] = \mu$  and finite variance  $\text{Var} [X_i] = \sigma^2$ , letting  $S_n = X_1 + X_2 + \dots + X_n$ , for any  $\varepsilon > 0$ , we have

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right) = 1 \quad (27)$$

### 2.4 Generalized Central Limit Theorem

For a sequence of iid r.v.s  $X_1, X_2, \dots, X_n$ , where  $n$  is the sample size, with finite mean  $E[X_i] = \mu$  and finite variance  $\text{Var} [X_i] = \sigma^2$ , letting  $S_n = X_1 + X_2 + \dots + X_n$ , we have

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) = \Phi(b) - \Phi(a) \quad (28)$$

More practically, we use

$$P(S \geq k) = P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \geq \frac{k - n\mu}{\sqrt{n}\sigma}\right) = 1 - \Phi\left(\frac{k - n\mu}{\sqrt{n}\sigma}\right) \quad (29)$$

### 3 Conditional distribution

A combination of conditional probability and marginal distribution.

#### 3.1 Discrete conditional distribution

recall the conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ for } P(B) > 0 \quad (30)$$

When A is now a r.v., we have the conditional distribution

$$p_{X|B}(k) = P(X = k|B) = \frac{P(\{X = k\} \cap B)}{P(B)} \quad (31)$$

##### 3.1.1 Conditional expectation of X, given event B

$$E[X|B] = \sum_k k P(X = k|B) \quad (32)$$

##### 3.1.2 Unconditiond pmf of X

$$p_X(k) = \sum_{i=1}^n p_{X|B_i}(k) P(B_i) \quad (33)$$

- From Equation 32 and Equation 33 we get

$$E[X] = \sum_{i=1}^n E[X|B_i] P(B_i) \quad (34)$$

##### 3.1.3 Conditioning on r.v.

When both X and Y are r.v.s, we can have the following two-variable function

$$p_{X|Y}(k|j) = P(X = k|Y = j) = \frac{P(\{X = k\}, \{Y = j\})}{P(Y = j)} = \frac{p_{X,Y}(k, j)}{p_Y(j)} \quad (35)$$

##### 3.1.4 Conditional expectation of X, given Y=Y

$$E[X|Y = j] = \sum_k k P(X = k|Y = j) = \sum_k k p_{X|Y}(k|j) \quad (36)$$

##### 3.1.5 Unconditioned pmf with 2 r.v.s

$$p_X(k) = \sum_j p_{X|Y}(k|j) p_Y(j) \quad (37)$$

- From this, we can derive the unconditioned expectation of X and Y

$$E[X] = \sum_k E[X|Y = j] p_Y(j) \quad (38)$$

##### 3.1.6 Joint pmf with 2 r.v.s

$$p_{X,Y}(k, j) = p_{X|Y}(k|j) p_Y(j) = p_{Y|X}(j|k) p_X(k) \quad (39)$$

### 3.2 Continuous conditional distribution

For continuous r.v.s, with both X, Y random variables, we have the conditional pdf of X given Y = y as

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad (40)$$

### 3.2.1 Conditional probability and expectation

$$P(X \in A|Y = y) = \int_A f_{X|Y}(t|y) dt \quad (41)$$

The conditional expectation of  $g(X)$

$$E[g(X)|Y = y] = \int_{-\infty}^{\infty} g(t) f_{X|Y}(t|y) dt \quad (42)$$

### 3.2.2 The unconditioned pdf and expectation of X

Given the conditional pdf  $f_{X|Y}(x|y)$ , we can derive the unconditioned pdf of X as

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) dy \quad (43)$$

$$E[g(X)] = \int_{-\infty}^{\infty} E[g(X)|Y = y] f_Y(y) dy \quad (44)$$

## 3.3 Conditional expectation

### 3.3.1 conditional expectation as a r.v.

Let X and Y jointly continuous r.v., The conditional expectation of X given Y is a new random variable dependent on Y  $v(Y)$

$$v(Y) = E[X|Y = y] \quad (45)$$

### 3.3.2 Conditioning and independence

recall that

- Discrete r.v. two discrete r.v.s are only independent iff

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \quad (46)$$

- Continuous r.v. two continuous r.v.s are only independent iff

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad (47)$$

now, If given pmf or pdf of X given Y, we now have the joint pmf

$$p_{X,Y}(x,y) = p_{X|Y}(x|y)p_Y(y) \quad (48)$$

and the joint pdf

$$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) \quad (49)$$

### 3.3.3 Independency of X and Y

discrete r.v. X and Y are independent iff

$$p_{X|Y}(x|y) = p_X(x) \quad (50)$$

continuous r.v. X and Y are independent iff

$$f_{X|Y}(x|y) = f_X(x) \quad (51)$$

### 3.4 Conditioning on the random variable

#### 3.4.1 Conditioning X on y

for independent r.v. X and Y, we can condition on Y and have the conditional expectation of X given Y = y as

$$E[g(X)|Y = y] = E[g(X)] \quad \text{and} \quad E[g(X)|Y = y] = E[g(X)] \quad (52)$$

#### 3.4.2 Conditioning X on X

For a r.v. X, we can condition on X itself, and have the conditional expectation of X given X = x as

$$E[g(X)|X = x] = g(x) \quad (53)$$