Module 4

Limit theorems

Topics: Sum of random variables, Correlation and independence, Law of large numbers, Convolution, Central limit theorem, Other functions

Sum of random variables

Consider $g(X,Y)$ to be a real-valued function of random vector (X,Y), e.g., $X+Y$, XY, X/Y (provided $Y \neq 0$), etc. Then $g(X,Y)$ is a random variable. We can obtain its distribution as: $F_{g(X,Y)}(z) = P(g(X,Y) \leq z) = P_{(X,Y)(\Omega)}(\{(x,y): g(x,y) \leq z\})$ for all $z \in \mathbb{R}$. Depending on the nature of $g(X,Y)$, we can obtain its mass/density function as well. In this module, we will focus on the sum function. This function arises in statistics quite often.

We begin with calculating $E[X+Y]$. For this we do not need to obtain mass/density function of $X+Y$. The law of unconscious statistician allows us to calculate $E[X+Y]$ directly from the joint mass/density function of (X,Y). For the discrete case,

$$\begin{aligned}
E[X+Y] = & \sum_{(x,y)\in(X,Y)(\Omega)} (x+y)\cdot p_{X,Y}(x,y) = \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} (x+y)\cdot p_{X,Y}(x,y) \\
& = \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} x\cdot p_{X,Y}(x,y) + \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} y\cdot p_{X,Y}(x,y) \\
& = \sum_{x\in X(\Omega)} x\cdot \sum_{y\in Y(\Omega)} p_{X,Y}(x,y) + \sum_{y\in Y(\Omega)} y\cdot \sum_{x\in X(\Omega)} p_{X,Y}(x,y) \\
& = \sum_{x\in X(\Omega)} x\cdot p_X(x) + \sum_{y\in Y(\Omega)} y\cdot p_Y(y) = E[X]+E[Y]
\end{aligned}$$

Note that the above derivation does not require independence of X,Y. *Show that the above result holds for the continuous case as well*. If we are interested in obtaining $E[X+Y+Z]$, we can consider $X+Y$ as a random variable, say Z'. Then with the above result, $E[X+Y+Z] = E[Z'+Z] = E[Z'] + E[Z] = E[X+Y] + E[Z] = E[X]+E[Y]+E[Z]$. We can extend the result to any number of random variables. In general, $E[a_0 + a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_0 + a_1E[X_1] + a_2E[X_2] + \cdots + a_nE[X_n]$, where $a_0, a_1, \ldots, a_n$ are constants.

Let us calculate $Var(X+Y)$. By definition, it is the expectation of squared distance of $X+Y$ from its mean. With the law of unconscious statistician,

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

Unlike expectation of sum, variance of sum is not the sum of variances. The additional term $E[(X-E[X])(Y-E[Y])]$ is known as the covariance of X,Y, and is denoted by $Cov(X,Y)$.

It measures the nature and degree of joint variation of X and Y with respect to their means. If increase/decrease in one random variable about its mean is (stochastically) accompanied by similar changes in the other, then $Cov(X,Y)$ is positive. On the other hand, if the changes are opposite, then $Cov(X,Y)$ is negative. If change in one does not significantly influence the other, then $Cov(X,Y)$ is near-zero. Sign of $Cov(X,Y)$ tells whether X,Y change in the same direction or in the opposite direction, and its magnitude indicates the strength of association between X,Y. The following examples illustrates all these cases.

Case-1a: $Cov(X,Y) = 1/6$

$p_{X,Y}(x,y)$	1	2	3	$p_X(x)$
0	1/6	1/12	1/12	1/3
1	1/12	1/6	1/12	1/3
2	1/12	1/12	1/6	1/3
$p_Y(y)$	1/3	1/3	1/3	1

Case-1b: $Cov(X,Y) = 4/9$

$p_{X,Y}(x,y)$	1	2	3	$p_X(x)$
0	2/9	1/9	0	1/3
1	1/9	1/9	1/9	1/3
2	0	1/9	2/9	1/3
$p_Y(y)$	1/3	1/3	1/3	1

Case-2a: $Cov(X,Y) = -1/6$

$p_{X,Y}(x,y)$	1	2	3	$p_X(x)$
0	1/12	1/12	1/6	1/3
1	1/12	1/6	1/12	1/3
2	1/6	1/12	1/12	1/3
$p_Y(y)$	1/3	1/3	1/3	1

Case-2b: $Cov(X,Y) = -4/9$

$p_{X,Y}(x,y)$	1	2	3	$p_X(x)$
0	0	1/9	2/9	1/3
1	1/9	1/9	1/9	1/3
2	2/9	1/9	0	1/3
$p_Y(y)$	1/3	1/3	1/3	1

Case-3: $Cov(X,Y) = 0$

$p_{X,Y}(x,y)$	1	2	3	$p_X(x)$
0	1/9	1/9	1/9	1/3
1	1/9	1/9	1/9	1/3
2	1/9	1/9	1/9	1/3
$p_Y(y)$	1/3	1/3	1/3	1

An alternate formula for covariance:

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])]$$
$$= E\big[XY - XE[Y] - E[X]Y - E[X]E[Y]\big]$$
$$= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y]$$
$$\Rightarrow Cov(X,Y) = E[XY] - E[X]E[Y]$$

In all the cases $E[X] = 1$ and $E[Y] = 2$. In case 1a and 1b, increase/decrease of one random variable about its mean is more likely to be accompanied by similar change in the other. So, covariance is positive in both cases. The magnitude is more in case 1b because the probability of similar change is higher than in case 1a. Case 2a and 2b are similar examples of negative covariance. In case 3, one does not seem to influence the other, and we have zero covariance. Note that covariance can be negative, unlike variance. Covariance of a random variable with itself is its variance; $Cov(X,X) = E[X^2] - E^2[X] = Var(X)$. Also, $Cov(aX + b, cY + d) = acCov(X,Y)$ for constants a, b, c, d. *Verify it yourself.*

We can obtain variance of sum of three or more variables using the formula for the two. Let $X + Y = Z'$. Then $Var(X + Y + Z) = Var(Z' + Z) = Var(Z') + Var(Z) + 2Cov(Z', Z) = Var(X + Y) + Var(Z) + 2Cov(X + Y, Z) = Var(X) + Var(Y) + Var(Z) + 2Cov(X,Y) + 2Cov(X + Y, Z)$. We can simplify the last term using the following result.

$$
\begin{aligned}
Cov(X_1+X_2+\cdots+X_n,Y) &= E[(X_1+X_2+\cdots+X_n)Y]-E[X_1+X_2+\cdots+X_n]E[Y] \\
&= E[X_1Y+X_2Y+\cdots+X_nY]-(E[X_1]+E[X_2]+\cdots+E[X_n])E[Y] \\
&= (E[X_1Y]-E[X_1]E[Y])+(E[X_2Y]-E[X_2]E[Y])+\cdots \\
&\quad +(E[X_nY]-E[X_n]E[Y]) = Cov(X_1,Y)+Cov(X_2,Y)+\cdots+Cov(X_n,Y)
\end{aligned}
$$

Using the above formula, $Var(X+Y+Z)=Var(X)+Var(Y)+Var(Z)+2Cov(X,Y)+2Cov(X,Z)+2Cov(Y,Z)$. We can extend this result to any number of random variables. In general, $Var(\sum_{i=1}^{n}X_i)=\sum_{i=1}^{n}Var(X_i)+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}Cov(X_i,X_j)$. *Verify this formula by the principle of mathematical induction*. Note that this formula is same as: $Var(\sum_{i=1}^{n}X_i)=\sum_{i=1}^{n}\sum_{j=1}^{n}Cov(X_i,X_j)$. Furthermore, $Var(\sum_{i=1}^{n}(a_iX_i+b_i))=\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_jCov(X_i,X_j)=\sum_{i=1}^{n}a_i^2Var(X_i)+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}a_ia_jCov(X_i,X_j)$ for constants a_i,b_i.

Correlation and independence

Covariance measures the nature and strength of association between two random variables. It may be used to tell whether X,Y are more strongly associated with one another than X,Z or not. Let us consider the stick breaking example of Module-3, where X and Y denote lengths of the 1st and 2nd pieces in meter. We obtained $f_{X,Y}(x,y)=2$ for $x,y\in(0,1)$ with $x+y<1$, and zero otherwise. Also, $f_X(t)=f_Y(t)=2(1-t)$ for $t\in(0,1)$, and zero otherwise.

$$\Rightarrow E[XY]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf_{X,Y}(x,y)dxdy=\int_0^1\int_0^{1-x}xy\cdot 2dydx=\int_0^1 x(1-x)^2dx=\frac{1}{12},$$

and $E[X]=E[Y]=\int_0^1 t\cdot 2(1-t)dt=\frac{1}{3}\Rightarrow Cov(X,Y)=E[XY]-E[X]E[Y]=-\frac{1}{36}$

A negative covariance is expected (*Why?*). Now, consider Z denote length of the 3rd piece in centimeter. We need to obtain $Cov(X,Z)$. Let Z' denote length of the 3rd piece in meter. Then (X,Y) and (X,Z') have identical distribution. Therefore, $Cov(X,Z')=Cov(X,Y)=-1/36$. Since $Z=100Z'$, then $Cov(X,Z)=Cov(X,100Z')=100Cov(X,Z')=-100/36$. If we ignore the units and compare $Cov(X,Y)$ and $Cov(X,Z)$, we will reach the wrong conclusion that X,Z are strongly associated than X,Y. Even if we are careful, the units may not convert into one another. For example, let X denote milage of a car, Y denote its speed, and Z denote the load. Here, we cannot compare $Cov(X,Y)$ and $Cov(X,Z)$ straightforwardly as the units of Y and Z do not convert into one another. In order to overcome these difficulties, we introduce a normalized measure of association, known as the correlation coefficient.

Correlation coefficient between random variables X,Y is defined as: $\rho_{X,Y}=Cov(X,Y)/\sigma_X\sigma_Y$, where σ_X, σ_Y are standard deviations of X, Y (i.e., σ_X^2, σ_Y^2 are variances of X, Y). Observe that $Cov(X,Y)$ and $\sigma_X\sigma_Y$ have same unit, and therefore, $\rho_{X,Y}$ is unitless. Moreover, $\rho_{X,Y}\in[-1,1]$, as shown below. This makes comparison between $\rho_{X,Y}$ and $\rho_{X',Y'}$ possible, where X,Y,X',Y' are arbitrary random variables. Observe that $\rho_{(aX+b),(cY+d)}=\rho_{X,Y}$.

$$Var\left(\frac{X}{\sigma_X} \pm \frac{X}{\sigma_Y}\right) = \frac{Var(X)}{\sigma_X^2} + \frac{Var(Y)}{(\pm\sigma_Y)^2} + 2\frac{Cov(X,Y)}{\sigma_X(\pm\sigma_Y)} = \frac{\sigma_X^2}{\sigma_X^2} + \frac{\sigma_Y^2}{\sigma_Y^2} \pm 2\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

$\Rightarrow 2 \pm 2\rho_{X,Y} = Var\left(\frac{X}{\sigma_X} \pm \frac{X}{\sigma_Y}\right) \geq 0 \Rightarrow 1 \pm \rho_{X,Y} \geq 0 \Rightarrow -1 \leq \rho_{X,Y} \leq 1$, as claimed.

Obtain $\rho_{X,Y}$ in the stick breaking example. Since $\rho_{X,Y}$ is a measure of association between X and Y, it should be zero whenever X and Y are independent. This, indeed, is the case. For the discrete case, using the law of unconscious statistician,

$$E[XY] = \sum_{(x,y)\in(X,Y)(\Omega)} xyp_{X,Y}(x,y) = \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} xyp_X(x)p_Y(y)\text{, due to independence}$$

$$= \sum_{x\in X(\Omega)} xp_X(x) \sum_{y\in Y(\Omega)} yp_Y(y) = \sum_{x\in X(\Omega)} xp_X(x)E[Y] = E[X]E[Y]$$

Then, $Cov(X,Y) = E[XY] - E[X]E[Y] = 0 \Rightarrow \rho_{X,Y} = 0$, as desired.

Show the above for the continuous case. Its converse, however, is not true, i.e., $\rho_{X,Y} = 0$ does not necessarily imply that X and Y are independent. Consider $X \sim U(-1,1)$ and $Y = X^2$. It is evident that X, Y are not independent. However, $Cov(X,Y) = E[XY] - E[X]E[Y] = E[X^3] - 0 \times E[X^2] = \int_{-1}^{1}(x^3/2)dx = 0 \Rightarrow \rho_{X,Y} = 0$. This is a shortcoming of correlation coefficient as a measure of association between random variables.

Law of large numbers

An electric car manufacturer wants to publicize kilometers run on a single charge. To get the number, it placed n cars on test in similar conditions. Due to uncontrollable parameters, the kilometers covered by these cars, denoted by $x_1, x_2, \dots, x_n$, can be different from one another. Then the average, i.e., $\bar{x} = \sum_{i=1}^{n} x_i/n$, is a good choice for kilometers run on a single charge. Before the test results $x_1, x_2, \dots, x_n$ are observed, these numbers are random variables. Let us denote these by $X_1, X_2, \dots, X_n$. Note that these are independent and identically distributed. The average $\bar{X} = \sum_{i=1}^{n} X_i/n$, too, is random now. This quantity arises in statistics very often. With the results obtained so far, we can determine mean and variance of $\bar{X}$. Let $E[X_1] = \mu$ and $Var(X_1) = \sigma^2$. *Show that $E[\bar{X}] = \mu$ and $Var(\bar{X}) = \sigma^2/n$.*

Note that $E[\bar{X}] = \mu$ does not change with n, but $Var(\bar{X}) = \sigma^2/n$ decreases as n increases, and eventually vanishes as $n \to \infty$. Variance of a random variable becoming zero implies its 'convergence' to its mean. Therefore, average of independent and identically distributed (*iid*) random variables 'converges' to the mean (of the random variables) as the sample size (i.e., the number of random variables considered in the average) goes to infinity. This observation is known as the law of large numbers.

Let $a_n = 1 - 1/n$ for $n = 1,2,3,\dots$. We know that a_n converges to 1 as $n \to \infty$. Here, the meaning of convergence is that the gap between a_n and 1 decreases and eventually vanishes as $n \to \infty$. Convergence of random variables is not that straightforward. Consider X_n and Y_n for $n = 1,2,3, \dots$ with mass functions: $p_X(1 - 1/\sqrt{n}) = p_X(1 + 1/\sqrt{n}) = 1/2$ and $p_Y(0) = 1/2n$, $p_Y(1) = 1 - 1/n$, $p_Y(2) = 1/2n$. Then $E[X_n] = E[Y_n] = 1$ and $Var(X_n) = Var(Y_n) = 1/n$ for all $n \geq 1$. Note that the mean does not change with n, but the variance decreases with n and eventually vanishes as $n \to \infty$. Therefore, X_n and Y_n both converges to 1 as $n \to \infty$. However, there is a difference between the convergence of X_n and Y_n. As n increases, all the values that X_n can take become increasingly closer to 1, whereas the probability that Y_n takes a value different from 1 becomes increasingly closer to 0. We must specify the nature of convergence when it involves random variables.

With the above observation in mind, we state the law of large numbers as follows:

Weak law: $\lim_{n\to\infty} P(|\bar{X} - \mu| > \epsilon) = 0$ for all $\epsilon > 0$

Strong law: $P\left(\lim_{n\to\infty} \bar{X} = \mu\right) = 1$

The weak law states that the probability of $\bar{X}$ deviating from μ (expected value of the random variables) vanishes as $n \to \infty$, whereas the strong law says that $\bar{X}$ becomes increasingly close to μ as $n \to \infty$. We can prove the weak law via Markov's and Chebyshev's Inequalities.

Markov's Inequality provides an upper bound to the probability of a non-negative random variable taking values larger than a threshold. For non-negative and continuous X, $E[X] = \int_0^\infty xf(x)dx = \int_0^t xf(x)dx + \int_t^\infty xf(x)dx \geq \int_t^\infty xf(x)dx \geq \int_t^\infty tf(x)dx = tP(X > t)$ for all $t > 0$, implying $P(X > t) \leq E[X]/t \;\; \forall t > 0$. *Show its validity for the discrete case.*

Chebyshev's Inequality provides an upper bound to the probability of any random variable taking values farther away from its mean than a threshold. For an arbitrary random variable X with mean μ and variance σ^2, let us define $Y = (X - \mu)^2$. Then Y is a non-negative random variable, and Markov's Inequality tells that $P(Y > t) \leq E[Y]/t \;\; \forall t > 0$. Note that $E[Y] = E[(X - \mu)^2] = Var(X) = \sigma^2$. Let us take $t = k^2\sigma^2$. Then for every $k > 0$, there is a $t > 0$. Then $P((X - \mu)^2 > k^2\sigma^2) \leq \sigma^2/k^2\sigma^2 \equiv P(|X - \mu| > k\sigma) \leq 1/k^2 \;\; \forall k > 0$.

Let us apply Chebyshev's Inequality to $\bar{X}$. Since $E[\bar{X}] = \mu$ and $Var(\bar{X}) = \sigma^2/n$, we have $P(|\bar{X} - \mu| > k\sigma/\sqrt{n}) \leq 1/k^2 \;\; \forall k > 0$. Let us take $k\sigma/\sqrt{n} = \epsilon$. Then for every $\epsilon > 0$, there is a $k > 0$ for finite n. Then $P(|\bar{X} - \mu| > \epsilon) \leq \sigma^2/n\epsilon^2 \;\; \forall \epsilon > 0$. As $n \to \infty$, the upper bound becomes zero for every positive ϵ. Since probability cannot be negative, the inequality holds as equality as $n \to \infty$. So, we have Then $\lim_{n\to\infty} P(|\bar{X} - \mu| > \epsilon) = 0 \; \forall \epsilon > 0$.

Law of large numbers ensures that the frequentist's measure converge to the true probability. Let p_A denote the true probability of an event A in a random experiment. p_A is unknown and the frequentist's measure says that it is $\lim_{n\to\infty} n_A/n$, where n_A denotes the number of times the event A occurs in n repetitions of the experiment. We can verify if $\lim_{n\to\infty} n_A/n = p_A$ or not. Let us define random variables $X_1, X_2, \dots, X_n$ as: $X_i = 1$ if i-th repetition of the random experiment yields the event A, else $X_i = 0$. Clearly, $\mu = p_A$, and by the law of large numbers, $\bar{X} = \sum_{i=1}^{n} X_i/n = n_A/n$ converges to p_A, as desired.

Convolution

We have obtained mean and variance of the sum of random variables without studying its distribution. Here, we obtain mass/density function of the sum of two random variables. Consider X and Y to be discrete and $Z = X + Y$. Then

$$\begin{aligned} p_Z(z) = P(Z = z) = P(X + Y = z) &= P\left(\bigcup_{x\in X(\Omega)} \{X = x, Y = z - x\}\right) \\ &= \sum_{x\in X(\Omega)} P(X = x, Y = z - x) = \sum_{x\in X(\Omega)} p_{X,Y}(x, z - x). \end{aligned}$$

Note that we can also write $p_Z(z) = \sum_{y\in Y(\Omega)} p_{X,Y}(z - y, y)$. These formulas are known as the convolution formula for the discrete case. For the continuous case,

$$\begin{aligned} F_Z(z) = P(Z \le z) = P(X + Y \le z) &= \int_{-\infty}^{\infty} P(X + Y \le z | X = x) P(X = x) \\ &= \int_{-\infty}^{\infty} P(Y \le z - x | X = x) f_X(x) dx = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} f_{Y|X=x}(y) dy\right) f_X(x) dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} \frac{f_{X,Y}(x, y)}{f_X(x)} f_X(x) dy dx = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{X,Y}(x, y) dy dx. \end{aligned}$$

Applying Leibnitz's integral rule, first to the outer integral and then to the inner integral,

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \int_{-\infty}^{\infty} \frac{d}{dz}\left(\int_{-\infty}^{z-x} f_{X,Y}(x, y) dy\right) dx = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x) dx.$$

Note that we can also write $f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(z - y, y) dy$. For sum of three or more random variables, we can apply the convolution formula recursively and obtain mass/density of the sum. *Consider the stick breaking example and obtain density function of* $X + Y$.

If the random variables X and Y are independent, the convolution formulas simplify.

$$p_Z(z) = \sum_{x\in X(\Omega)} p_X(x) p_Y(z - x) \ \text{ for the discrete case}$$

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx \ \text{ for the continuous case}$$

Applying the above formulas, we can obtain interesting results about some of the popular random variables. Let X and Y denote two independent binomial random variables with the same success probability. Let $X \sim Bin(m, p)$ and $Y \sim Bin(n, p)$. Then $Z = X + Y$ takes values $0,1,2, \dots, m + n$. For such a z, $p_Z(z)$ is obtained as follows:

$$p_Z(z) = \sum_{x \in X(\Omega)} p_X(x) p_Y(z-x) = \sum_{x=0}^{m} \binom{m}{x} p^x q^{m-x} p_Y(z-x), \text{where } q = 1-p.$$

We shall be careful about writing $p_Y(z-x) = \binom{n}{z-x} p^{z-x} q^{n-z+x}$, as it implicitly assumes $0 \le z-x \le n \equiv z-n \le x \le z$. Outside this range, $p_Y(z-x) = 0$. We can ensure that we are within this range by considering $x = \max(0, z-n)$ to $\min(m, z)$. So,

$$p_Z(z) = \sum_{x=\max(0,z-n)}^{min(m,z)} \binom{m}{x} p^x q^{m-x} \binom{n}{z-x} p^{z-x} q^{n-z+x} = \sum_{x=\max(0,z-n)}^{min(m,z)} \binom{m}{x} \binom{n}{z-x} p^z q^{m+n-z}$$
$$= \binom{m+n}{z} p^z q^{m+n-z} \sum_{x=\max(0,z-n)}^{min(m,z)} \binom{m}{x} \binom{n}{z-x} \Big/ \binom{m+n}{z}$$

The above sum is 1, as explained next. Then $p_Z(z) = \binom{m+n}{z} p^z q^{m+n-z}$ for $z = 0,1, \dots, m+n$. So, the sum of two independent binomial random variables with the same success probability is another binomial random variable with the same success probability. This result can be extended to more than two binomial random variables. A similar result holds for independent Poisson random variables. *If $X \sim Pois(\lambda, s)$ and $Y \sim (\lambda, t)$ are independent, then show that $X + Y \sim Pois(\lambda, s + t)$.* In some books μ_1 and μ_2 are used in places of λs and λt. Then the result is: If $X \sim Pois(\mu_1)$ and $Y \sim Pois(\mu_2)$ are independent, then $X + Y \sim Pois(\mu_1 + \mu_2)$. Again, this result can be extended to more than two Poisson random variables.

Consider m white and n black balls are put together in an urn. We draw $z \in \{0,1, \dots, m+n\}$ number of balls from the urn randomly. Let X denote the number of white balls drawn. It's a discrete random variable that takes values in $\{\max(0, z-n), \max(0, z-n)+1, \max(0, z-n)+2, \dots, min(m, z)\}$. For x in this range,

$$p_X(x) = P(\text{Drawing } x \text{ white balls from a pool of } m \text{ white and } n \text{ black balls})$$
$$= \binom{m}{x} \binom{n}{z-x} \Big/ \binom{m+n}{z}, \text{by classical measure.}$$

The above distribution is known as the hyper-geometric distribution. It also ensures that the sum of the terms $\binom{m}{x}\binom{n}{z-x}/\binom{m+n}{z}$ over $x = \max(0, z-n)$ to $\min(m, z)$ is 1, which is used in obtaining the earlier result about binomial random variables.

Let us consider the case of normal random variables. Let $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ to be independent normal random variables. By the convolution formula,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{\{(z-x)-\mu_2\}^2}{2\sigma_2^2}} dx$$
$$= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} e^{-\frac{\{z-(\mu_1+\mu_2)\}^2}{2(\sigma_1^2+\sigma_2^2)}} \int_{-\infty}^{\infty} \sqrt{\frac{\sigma_1^2+\sigma_2^2}{2\pi\sigma_1^2\sigma_2^2}} e^{-\frac{1}{2}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{\{(z-x)-\mu_2\}^2}{\sigma_2^2} - \frac{\{z-(\mu_1+\mu_2)\}^2}{\sigma_1^2+\sigma_2^2}\right]} dx$$

Let us write the integral as: $\int_{-\infty}^{\infty} e^{-g(x)/2}/\sqrt{2\pi v^2}\,dx$, where $v^2 = \sigma_1^2\sigma_2^2/(\sigma_1^2+\sigma_2^2)$ and

$$g(x) = \frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(z-x-\mu_2)^2}{\sigma_2^2} - \frac{(z-\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}$$
$$= x^2\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right) - 2x\left(\frac{\mu_1}{\sigma_1^2}+\frac{z-\mu_2}{\sigma_2^2}\right) + \left\{\frac{\mu_1^2}{\sigma_1^2}+\frac{(z-\mu_2)^2}{\sigma_2^2}-\frac{(z-\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}\right\}$$

$$= \frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}\left[x^2 - 2x\frac{\mu_1\sigma_2^2+(z-\mu_2)\sigma_1^2}{\sigma_1^2+\sigma_2^2} + \frac{\mu_1^2\sigma_2^2(\sigma_1^2+\sigma_2^2)+(z-\mu_2)^2\sigma_1^2(\sigma_1^2+\sigma_2^2)-\{(z-\mu_2)-\mu_1\}^2\sigma_1^2\sigma_2^2}{(\sigma_1^2+\sigma_2^2)^2}\right]$$
$$= \frac{1}{v^2}\left[x^2 - 2x\frac{\mu_1\sigma_2^2+(z-\mu_2)\sigma_1^2}{\sigma_1^2+\sigma_2^2} + \frac{\mu_1^2\sigma_2^4+(z-\mu_2)^2\sigma_1^4+2(\mu_1\sigma_2^2)\{(z-\mu_2)\sigma_1^2\}}{(\sigma_1^2+\sigma_2^2)^2}\right]$$
$$= \frac{(x-u)^2}{v^2}, \text{where } u = \frac{\mu_1\sigma_2^2+(z-\mu_2)\sigma_1^2}{\sigma_1^2+\sigma_2^2} \text{ (a constant, given } z).$$

Then the integral becomes: $\int_{-\infty}^{\infty} e^{-(x-u)^2/2v^2}/\sqrt{2\pi v^2}\,dx = 1$, as the integrand is the density function of $N(u, v^2)$. Then $f_Z(z) = e^{-\{z-(\mu_1+\mu_2)\}^2/2(\sigma_1^2+\sigma_2^2)}\big/\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}$. Clearly, sum of independent $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ is $N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$. We can extend this result to multiple independent normal random variables. Earlier, we noted that $X \sim N(\mu, \sigma^2)$ implies $aX+b \sim N(a\mu+b, a^2\sigma^2)$. Then $a_0 + a_1X_1 + a_2X_2 + \cdots + a_nX_n$, where $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1,2,\ldots,n$ are independent, is $N(a_0 + \sum_{i=1}^n a_i\mu_i, \sum_{i=1}^n a_i^2\sigma_i^2)$.

Central limit theorem

Let $X_1, X_2, \ldots, X_n$ denote *iid* random variables with expectation μ and variance σ^2. Let $\bar{X} = \sum_{i=1}^n X_i/n$ denote their average. We established that $\bar{X}$ converges to μ as $n \to \infty$; specifically, the probability of $\bar{X}$ deviating from μ by any positive amount is zero as $n \to \infty$. This does not mean $\bar{X}$ takes only one value μ as $n \to \infty$. Here, we will try to obtain the distribution of $\bar{X}$ as $n \to \infty$. This will require the use of moment generating function.

Moment generating function of random variable X with respect to real-valued parameter s is: $m_X(s) = E[e^{sX}]$ provided the expectation exists. It helps us obtain moments of X. If $m_X(s)$ is defined in an interval around $s = 0$, then k-th moment of X for $k = 1,2,3,\ldots$ is same as the k-th derivative of $m_X(s)$ at $s = 0$. Its true usefulness is in comparing random variables. If two random variables have the same moment generating function in an interval around $s = 0$, then they have the same distribution. Let us see some use of this property.

Consider $X \sim Bin(m,p)$ and $Y \sim Bin(n,p)$ to be independent. Earlier, we showed that $X + Y \sim Bin(m+n, p)$. We can do the same with moment generating function.

$$m_X(s) = E[e^{sX}] = \sum_{x=0}^{m} e^{sx}\binom{m}{x}p^x q^{m-x} = \sum_{x=0}^{m}\binom{m}{x}(pe^s)^x q^{m-x} = (pe^s + q)^m \ \forall s \in \mathbb{R}$$

by the binomial theorem. In a similar manner, $m_Y(s) = (pe^s + q)^n \ \forall s \in \mathbb{R}$. Now,

$$m_{X+Y}(s) = E\left[e^{s(X+Y)}\right] = E[e^{sX}e^{sY}] = E[e^{sX}]E[e^{sY}], \text{as } e^{sX}, e^{sY} \text{ are independent}$$
$$= m_X(s)m_Y(s) = (pe^s + q)^{m+n} \ \forall s \in \mathbb{R}$$

Clearly, moment generating function of $X + Y$ is same as that of $Bin(m + n, p)$, and both are defined in an interval around $s = 0$. Thus, $X + Y \sim Bin(m + n, p)$. *Prove the similar result about Poisson random variable using moment generating function.*

Let us consider $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ to be independent. We showed that $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. We can do the same with moment generating function.

$$m_X(s) = E[e^{sX}] = \int_{-\infty}^{\infty} e^{sx}\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}dx = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{x^2-2x(\mu_1+\sigma_1^2 s)+\mu_1^2}{2\sigma_1^2}}dx$$
$$= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{\{x-(\mu_1+\sigma_1^2 s)\}^2-2\mu_1\sigma_1^2 s-\sigma_1^4 s^2}{2\sigma_1^2}}dx = e^{\mu_1 s+\frac{1}{2}\sigma_1^2 s^2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{\{x-(\mu_1+\sigma_1^2 s)\}^2}{2\sigma_1^2}}dx$$
$$= e^{\mu_1 s+\frac{1}{2}\sigma_1^2 s^2} \ \forall s \in \mathbb{R}, \text{as the integrand is density function of } N(\mu_1 + \sigma_1^2 s, \sigma_1^2).$$

In a similar manner, $m_Y(s) = e^{\mu_2 s+\frac{1}{2}\sigma_2^2 s^2} \ \forall s \in \mathbb{R}$. Since X, Y are independent, $m_{X+Y}(s) = m_X(s)m_Y(s)$, as observed earlier. Then $m_{X+Y}(s) = e^{(\mu_1+\mu_2)s+\frac{1}{2}(\sigma_1^2+\sigma_2^2)s^2} \ \forall s \in \mathbb{R}$. Clearly, moment generating function of $X + Y$ is same as that of $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, and both are defined in an interval around $s = 0$. Thus, $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. *Now, prove that* $aX + bY + c \sim N(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2)$ *for independent* $X \sim N(\mu_1, \sigma_1^2)$ *and* $Y \sim N(\mu_2, \sigma_2^2)$ *and constants* a, b, c *using moment generating function.*

Now, let us focus on $\bar{X} = \sum_{i=1}^{n} X_i/n$, where $X_1, X_2, \dots, X_n$ are *iid* random variables with mean μ and variance σ^2. Earlier, we found that $E[\bar{X}] = \mu$ and $Var(\bar{X}) = \sigma^2/n$. Let us standardize $\bar{X}$, i.e., construct $\bar{Z} = (\bar{X} - \mu)/(\sigma/\sqrt{n})$. Note that $E[\bar{Z}] = 0$ and $Var(\bar{Z}) = 1$. Standardizing a random variable involves subtracting its mean and then dividing by its standard deviation. *Verify that mean and variance of any standardized random variable are* 0 *and* 1. The central limit theorem says that $\bar{Z}$ converges with $N(0,1)$ as $n \to \infty$. Formally,

$$\lim_{n\to\infty} P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le z\right) = \Phi(z) \ \forall z \in \mathbb{R}$$

where Φ denotes distribution function of the standard normal random variable and $\bar{X}$ denotes the average of *iid* random variables $X_1, X_2, \dots, X_n$ with mean μ and variance σ^2. Note that $X_1, X_2, \dots, X_n$ can be discrete as well as continuous. This tells how normal distribution arises. Whenever the random variable of interest can be thought of as average/sum of a large number of *iid* random variables, then the random variable of interest follows normal distribution. For the same reason, we can approximate $Bin(n,p)$ for large n by $N(np,npq)$ and $Pois(\lambda,t)$ for large t by $N(\lambda t,\lambda t)$. *Do you see why*?

Proof of the central limit theorem involves establishing that the moment generating function of $\bar{Z}$ with parameter $s \in \mathbb{R}$ converges to $e^{s^2/2}$ as $n \to \infty$, which is same as that of $N(0,1)$. Let us define $Z_i = (X_i - \mu)/\sigma$ for $i = 1,2,\ldots,n$. Then $Z_1, Z_2, \ldots, Z_n$ are *iid* random variables and

$$\frac{\sum_{i=1}^n Z_i}{\sqrt{n}} = \frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n}\sigma} = \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu) = \bar{Z}$$

$$\begin{aligned}\Rightarrow m_{\bar{Z}}(s) &= E\left[e^{s\bar{Z}}\right] = E\left[e^{s(Z_1+Z_2+\cdots+Z_n)/\sqrt{n}}\right] = E\left[e^{(s/\sqrt{n})Z_1}e^{(s/\sqrt{n})Z_2}\cdots e^{(s/\sqrt{n})Z_n}\right] \\ &= E\left[e^{sZ_1/\sqrt{n}}\right]E\left[e^{sZ_2/\sqrt{n}}\right]\cdots E\left[e^{sZ_n/\sqrt{n}}\right], \text{due to independence} \\ &= E^n\left[e^{sZ_1/\sqrt{n}}\right], \text{due to identicalness}\end{aligned}$$

$$\begin{aligned}\text{Now, } E\left[e^{sZ_1/\sqrt{n}}\right] &= E\left[1 + sZ_1/\sqrt{n} + \frac{\left(sZ_1/\sqrt{n}\right)^2}{2} + \frac{\left(sZ_1/\sqrt{n}\right)^3}{3!} + \frac{\left(sZ_1/\sqrt{n}\right)^4}{4!} + \cdots\right] \\ &= 1 + \frac{s}{\sqrt{n}}E[Z_1] + \frac{s^2}{2n}E[Z_1^2] + \frac{s^3}{3!\,n^{3/2}}E[Z_1^3] + \frac{s^4}{4!\,n^{4/2}}E[Z_1^4] + \cdots \\ &= 1 + \frac{1}{n}\left[\frac{s^2}{2} + \frac{s^3E[Z_1^3]}{3!\,n^{1/2}} + \frac{s^4E[Z_1^4]}{4!\,n^{2/2}} + \cdots\right], \text{as } E[Z_1] = 0 \text{ and } E[Z_1^2] = 1 \\ &= 1 + \frac{g(n)}{n}, \text{where } g(n) = \frac{s^2}{2} + \frac{s^3E[Z_1^3]}{3!\,n^{1/2}} + \frac{s^4E[Z_1^4]}{4!\,n^{2/2}} + \cdots\end{aligned}$$

$E[Z_1^k] = E[\{(X_1 - \mu)/\sigma\}^k] = E[(X_1 - \mu)^k]/\sigma^k$ for $k = 1,2,3,\ldots$ are known as standardized moments of X_1. $E[Z_1^1] = 0$ and $E[Z_1^2] = 1$, as noted earlier. It can be proved that $E[Z_1^k]$ is finite for all k. Then $g(n) \to s^2/2$ as $n \to \infty$.

$$\Rightarrow \lim_{n\to\infty} m_{\bar{Z}}(s) = \lim_{n\to\infty}\left(1 + \frac{g(n)}{n}\right)^n = \lim_{n\to\infty}\left(1 + \frac{s^2/2}{n}\right)^n = e^{s^2/2}, \text{as desired.}$$

Other functions

Among other real-valued functions of a random vector, the maximum and minimum function are important. Let us consider $Y = \max\{X_1, X_2, \ldots, X_n\}$, where $X_1, X_2, \ldots, X_n$ are *iid* random variables with distribution function F_X. Then

$$\begin{aligned}F_Y(y) = P(Y \le y) &= P(\max\{X_1, X_2, \ldots, X_n\} \le y) = P(X_1 \le y, X_2 \le y, \ldots, X_n \le y) \\ &= P(X_1 \le y)P(X_2 \le y)\cdots P(X_n \le y) = F_X^n(y) \text{ for all } y \in \mathbb{R}.\end{aligned}$$

In a similar manner, we can study $Z = \min\{X_1, X_2, \ldots, X_n\}$.

$$\begin{aligned}F_Z(z) = P(Z \le z) &= 1 - P(Z > z) = 1 - P(\min\{X_1, X_2, \ldots, X_n\} > z) \\ &= 1 - P(X_1 > z, X_2 > z, \ldots, X_n > z) \\ &= 1 - P(X_1 > z)P(X_2 > z)\cdots P(X_n > z) = 1 - \{1 - F_X(z)\}^n \text{ for all } z \in \mathbb{R}.\end{aligned}$$

If $X_1, X_2, \ldots, X_n$ are continuous and f_X is their density function, then

$$f_Y(y) = \frac{d}{dy} F_X^n(y) = nF_X^{n-1} \frac{dF_X(y)}{dy} = nF_X^{n-1} f_X(y) \text{ for all } y \in \mathbb{R}.$$

$$\begin{aligned} f_Z(z) &= \frac{d}{dz}(1 - \{1 - F_X(z)\}^n) = -n\{1 - F_X(z)\}^{n-1} \frac{d\{1 - F_X(z)\}}{dz} \\ &= n\{1 - F_X(z)\}^{n-1} f_X(z) \text{ for all } z \in \mathbb{R}. \end{aligned}$$

Another function that we encounter sometimes is the product function. Let $Z = XY$, where X and Y are independent continuous random variables. Then

$$\begin{aligned} F_Z(z) = P(Z \le z) = P(XY \le z) &= \int_{-\infty}^{\infty} P(XY \le z | X = x) P(X = x) \\ &= \int_{-\infty}^{0} P(xY \le z | X = x) f_X(x) dx + \int_{0}^{\infty} P(xY \le z | X = x) f_X(x) dx \\ &= \int_{-\infty}^{0} P(Y \ge z/x) f_X(x) dx + \int_{0}^{\infty} P(Y \le z/x) f_X(x) dx \\ &= \int_{-\infty}^{0} \{1 - F_Y(z/x)\} f_X(x) dx + \int_{0}^{\infty} F_Y(z/x) f_X(x) dx \text{ for all } z \in \mathbb{R}. \end{aligned}$$

Using Applying Leibnitz's integral rule,

$$\begin{aligned} f_Z(z) = \frac{dF_Z(z)}{dz} &= \int_{-\infty}^{0} \left\{-f_Y(z/x)\frac{1}{x}\right\} f_X(x) dx + \int_{0}^{\infty} \left\{f_Y(z/x)\frac{1}{x}\right\} f_X(x) dx \\ &= \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(z/x) dx \text{ for all } z \in \mathbb{R}. \end{aligned}$$

If we are interested in $Z = X/Y$, then following the above approach,

$$\begin{aligned} F_Z(z) = P(Z \le z) = P(X/Y \le z) &= \int_{-\infty}^{\infty} P(X/Y \le z | Y = y) P(Y = y) \\ &= \int_{-\infty}^{0} P(X/y \le z | Y = y) f_Y(y) dy + \int_{0}^{\infty} P(X/y \le z | Y = y) f_Y(y) dy \\ &= \int_{-\infty}^{0} P(X \ge yz) f_Y(y) dy + \int_{0}^{\infty} P(X \le yz) f_Y(y) dy \\ &= \int_{-\infty}^{0} \{1 - F_X(yz)\} f_Y(y) dy + \int_{0}^{\infty} F_X(yz) f_Y(y) dy \text{ for all } z \in \mathbb{R}. \end{aligned}$$

Using Applying Leibnitz's integral rule,

$$\begin{aligned} f_Z(z) = \frac{dF_Z(z)}{dz} &= \int_{-\infty}^{0} \{-f_X(yz) y\} f_Y(y) dy + \int_{0}^{\infty} \{f_X(yz) y\} f_Y(y) dy \\ &= \int_{-\infty}^{\infty} |y| f_X(yz) f_Y(y) dy \text{ for all } z \in \mathbb{R}. \end{aligned}$$

Practice problems

Book-1: A Modern Introduction to Probability and Statistics by Dekking et al.

Mean and variance of the sum
Book-1, Chapter-10, Exercise No. 10, 13, 16, 18, 20

Law of large numbers
Book-1, Chapter-13, Exercise No. 2, 3, 5, 12

Distribution of the sum
Book-1, Chapter-11, Exercise No. 1, 5, 6, 11

Central limit theorem
Book-1, Chapter-14, Exercise No. 2, 3, 6, 8, 9