7. Properties of Expectation

Theorem If $a \le X \le b$, then $a \le E(X) \le b$.

Proof

We will prove the discrete case. First note that

$$E(X) = \sum xp(x) \ge \sum ap(x) = a.$$

In the same manner,

$$E(X) = \sum xp(x) \le \sum bp(x) = b.$$

The proof in the continuous case is similar.

7.1. Expectation of Sums of Random Variables

Theorem

(a) If X and Y are jointly discrete with joint probability mass function $p_{X,Y}$, then

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

(b) If X and Y are jointly continuous with joint probability density function $f_{X,Y}$, then

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) \, \mathrm{d}x \, \mathrm{d}y .$$

Proof is simple and skipped.

7.1. Expectation of Sums of Random Variables

Example An accident occurs at a point X that is uniformly distributed on a road of length L. At the time of the accident an ambulance is at a location Y that is also uniformly distributed on the road. Assuming that X and Y are independent, find the expected distance between the ambulance and the point of the accident.

Solution We need to compute $E[|X-Y|]$. Since the joint density function of X and Y is

$$f(x,y) = \frac{1}{L^2}, \quad 0 < x, y < L,$$

we obtain $$E[|X-Y|] = \frac{1}{L^2}\int_0^L \int_0^L |x-y| \, \mathrm{dy}\, \mathrm{dx}.$$

Now, $$\begin{aligned}\int_0^L |x-y| \, \mathrm{dy} &= \int_0^x (x-y)\, \mathrm{dy} + \int_x^L (y-x)\, \mathrm{dy} \\ &= \frac{x^2}{2} + \frac{L^2}{2} - \frac{x^2}{2} - x(L-x) \\ &= \frac{L^2}{2} + x^2 - xL.\end{aligned}$$

Therefore,
$$E[|X-Y|] = \frac{1}{L^2}\int_0^L \left(\frac{L^2}{2} + x^2 - xL\right) \mathrm{dx} = \frac{L}{3}.$$

7.1. Expectation of Sums of Random Variables

Remarks

(1) If $g(x,y) \geq 0$ whenever $p_{X,Y}(x,y) > 0$, then $E[g(X,Y)] \geq 0$.

(2) $E[g(X,Y)+h(X,Y)] = E[g(X,Y)] + E[h(X,Y)]$.

(3) $E[g(X)+h(Y)] = E[g(X)] + E[h(Y)]$.

(4) **Monotone Property**
If jointly distributed random variables X and Y satisfy $X \leq Y$. Then,

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

Proof of (4)

From (1), consider

$$E(Y) - E(X) = E[Y - X] \geq 0.$$

7.1. Expectation of Sums of Random Variables

Important special case

(Mean of sum = sum of means)

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

This of course can be easily extended to

$$E(a_1X_1 + \cdots + a_nX_n) = a_1E(X_1) + \cdots + a_nE(X_n).$$

They hold no matter X and Y are dependent or independent.

7.1. Expectation of Sums of Random Variables

Example (Sample Mean)

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having distribution function F and expected value μ. Such a sequence of random variables is said to constitute a sample from the distribution F. Define the sample mean $\overline{X}$, as follows:

$$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k. \qquad \text{Find } E(\bar{X}).$$

Solution

$$\begin{aligned} E(\overline{X}) &= E\left[\frac{1}{n}\sum_{k=1}^{n} X_k\right] \\ &= \frac{1}{n}\sum_{k=1}^{n} E(X_k) \quad = \quad \frac{1}{n}\sum_{k=1}^{n} \mu \quad = \quad \mu. \end{aligned}$$

That is, the expected value of the sample mean is μ, the mean of the distribution. When the distribution mean μ is unknown, the sample mean is often used in statistics to estimate it.

7.1. Expectation of Sums of Random Variables

Example (Boole's Inequality)

Let $A_1, \ldots, A_n$ be events in a probability space. Define the indicator variables I_k, $k = 1, \ldots, n$ by

$$I_k = \begin{cases} 1, & \text{if } A_k \text{ occurs} \\ 0, & \text{otherwise} \end{cases}.$$

Let $X = \sum_{k=1}^{n} I_k$ and $Y = \begin{cases} 1, & \text{if one of } A_k \text{ occurs} \\ 0, & \text{otherwise} \end{cases} = \begin{cases} 1, & \text{if } \cup_{k=1}^{n} A_k \text{ occurs} \\ 0, & \text{otherwise} \end{cases} = I_{\cup_{k=1}^{n} A_k}.$

We have

$$E(X) = \sum_{k=1}^{n} E(I_k) = \sum_{k=1}^{n} P(I_k = 1) = \sum_{k=1}^{n} P(A_k),$$

$$E(Y) = E\left[I_{\cup_{k=1}^{n} A_k}\right] = P\left(I_{\cup_{k=1}^{n} A_k} = 1\right) = P\left(\cup_{k=1}^{n} A_k\right).$$

Furthermore, it is easy to see that $Y \leq X$, therefore $E(Y) \leq E(X)$.

And this is equivalent to $P\left(\cup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} P(A_k)$, commonly known as **Boole's Inequality**.

7.1. Expectation of Sums of Random Variables

Example (Mean of Binomial)

Let $X \sim$Binomial(n, p) and

$$I_k = \begin{cases} 1, & \text{if the kth trial is a success} \\ 0, & \text{if the kth trial is a failure} \end{cases}.$$

Important observation:

$$X = I_1 + I_2 + \cdots + I_n.$$

Then,

$$E(X) = E(I_1) + \cdots + E(I_n) = p + \cdots + p = np.$$

7.1. Expectation of Sums of Random Variables

Example (Mean of Negative Binomial)

If independent trials, having a constant probability p of being successes, are performed, determine the expected number of trials required to amass a total of r successes.

Solution If X denotes the number of trials needed to amass a total of r successes, then X is a negative binomial random variable. It can be represented by

$$X = X_1 + X_2 + \cdots + X_r,$$

where X_1 is the number of trials required to obtain the first success, X_2 the number of additional trials until the second success is obtained, X_3 the number of additional trials until the third success is obtained, and so on. That is, X_i represents the number of additional trials required, after the $(i-1)$st success, until a total of i successes is amassed. A little thought reveals that each of the random variable is a geometric random variable with parameter p. Hence, $E[X_i] = 1/p$, $i = 1, 2, \ldots, r$; and thus

$$E(X) = E(X_1) + \cdots + E(X_r) = \frac{r}{p}.$$

7.1. Expectation of Sums of Random Variables

Example (Mean of Hypergeometric)

If n balls are randomly selected from an urn containing N balls of which m are white, find the expected number of white balls selected.

Solution Let X denote the number of white balls selected, and represent X as

$$X = X_1 + \cdots + X_m$$

$$P(X=x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$$

where

$$X_i = \begin{cases} 1, & \text{if the } i\text{th white ball is selected} \\ 0, & \text{otherwise} \end{cases}.$$

已知 ith ball 被选，选其他 n-1 个就可以了.

Now,

$$\begin{aligned} E(X_i) &= P(X_i = 1) \\ &= P(i\text{th white ball is selected}) \\ &= \frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{n}} \end{aligned}$$

$$\rightarrow \frac{\frac{(N-1)!}{(n-1)!\,\cancel{(N-n)!}}}{\frac{N!}{n!\,\cancel{(N-n)!}}} = \frac{n}{N}$$

Hence

$$E(X) = E(X_1) + \cdots + E(X_m) = \frac{nm}{N}.$$

7.1. Expectation of Sums of Random Variables

Example (Expected Number of Matches)

A group of n people throw their hats into the center of a room. The hats are mixed up, each person randomly selects one. Find the expected number of people who get back their own hats.

Solution For $1 \leq k \leq n$, define $I_k = \begin{cases} 1, & \text{if the kth person gets back his hat} \\ 0, & \text{otherwise} \end{cases}$.

Then $X := I_1 + I_2 + \cdots + I_n$

denotes the number of people who get back their hats. Now

$$P(I_k = 1) = \frac{1}{n}.$$

我淨係care k-th person 拿不拿到，
其他人能否拿到與我無關。

$\therefore P(I_k = 1) = \frac{(n-1)!}{n!} = \frac{1}{n}$

Then
$$\begin{aligned} E(X) &= E[I_1 + I_2 + \cdots + I_n] \\ &= P(I_1 = 1) + P(I_2 = 1) + \cdots + P(I_n = 1) \\ &= \frac{1}{n} + \cdots + \frac{1}{n} \\ &= 1. \end{aligned}$$

7.1. Expectation of Sums of Random Variables

Example (Coupon-collecting)

Suppose that there are N different types of coupons and each time one obtains a coupon it is equally likely to be any one of the N types.

(a) Find the expected number of different types of coupons that are contained in a set of n coupons. ← 即 AABBCD if $n=6$.

(b) Find the expected number of coupons one need amass before obtaining a complete set of at least one of each type.

Solution

(a) Let X denote the number of different types of coupons in the set of n coupons. We compute $E[X]$ by using the representation $X = X_1 + \cdots + X_N$

where $X_i = \begin{cases} 1, & \text{if at least one type } i \text{ coupon is contained in the set of } n \\ 0, & \text{otherwise} \end{cases}$. ← 有A即可, 不需 count 有多少A.

Now, $E[X_i] = P(X_i = 1)$

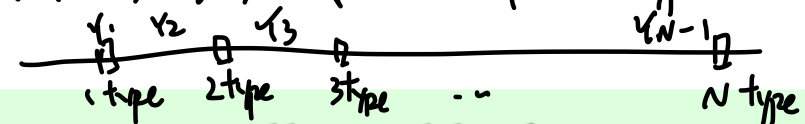
$= 1 - P(\text{no type } i \text{ coupons are contained in the set of } n) = 1 - \left(\frac{N-1}{N}\right)^n$.

Hence $E[X] = E[X_1] + \cdots + E[X_N] = N\left[1 - \left(\frac{N-1}{N}\right)^n\right]$.

7.1. Expectation of Sums of Random Variables

Solution (cont.)

Define Y as 我需要再抽幾多先有一個新嘅type出現.



(b) Let Y denote the number of coupons collected before a complete set is attained. We compute $E[Y]$ by using the same technique as we used in computing the mean of a negative binomial random variable. That is, define $Y_i, i = 0, 1, \ldots, N-1$ to be the number of additional coupons that need be obtained after i distinct types have been collected in order to obtain another distinct type, and note that

$$Y = Y_0 + Y_1 + \cdots + Y_{N-1}.$$

When i distinct types of coupons have already been collected, it follows that a new coupon obtained will be of a distinct type with probability $(N-i)/N$. Therefore,

因為有 $\frac{1}{N}$ 落入任何一type, 而現有 $(N-i)$ type 有可能係新

cond. prob.

$$P(Y_i = k) = \frac{N-i}{N}\boxed{\left(\frac{i}{N}\right)^{k-1}} \quad k \geq 1,$$

之前 k-1 張都在前 i 種 coupon 中輪迴

or in other words, Y_i is a geometric random variable with parameter $(N-i)/N$. Hence

$$E[Y_i] = \frac{N}{N-i}$$

因為 Y_i 是多次才到下一次成功(新type), 而成功prob是$\left(\frac{N-i}{N}\right)$.

implying that

$$E[Y] = 1 + \frac{N}{N-1} + \frac{N}{N-2} + \cdots + \frac{N}{1} = N\left[1 + \cdots + \frac{1}{N-1} + \frac{1}{N}\right]. \approx N \log N$$

7.2 Covariance, Variance of Sums, and Correlations

Definitions *The* ***covariance*** *of jointly distributed random variables X and Y, denoted by* $\operatorname{cov}(X,Y)$, *is defined by*

$$\operatorname{cov}(X,Y) = E(X-\mu_X)(Y-\mu_Y)$$

where μ_X, μ_Y denote the means of X and Y respectively.

Remark If $\operatorname{cov}(X,Y) \neq 0$, we say that X and Y are **correlated**.
If $\operatorname{cov}(X,Y) = 0$, we say that X and Y are **uncorrelated**.

Warning: Correlation does NOT imply causation.

7.2 Covariance, Variance of Sums, and Correlations

Example

- The Japanese eat little fat and suffer fewer heart attacks than the British or Americans.
- The French eat a lot of fat and also suffer fewer heart attacks than the British or Americans.
- The Italians drink a lot of red wine and also suffer fewer heart attacks than the British or Americans.

Conclusions:

- Eat and drink what you like. ***uncorrelated***
- Speaking English is apparently what kills you.

false causation by correlation

7.2 Covariance, Variance of Sums, and Correlations

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

Proof

Let μ_X, μ_Y denote respectively $E(X)$ and $E(Y)$.

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

此处的 $E[X]$ 就是 μ_X.

7.2 Covariance, Variance of Sums, and Correlations

Theorem *If X and Y are independent, then for any functions $g, h : \mathbb{R} \to \mathbb{R}$, we have*

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$

Proof Ross [p.322] proves this in the continuous case. Let's prove the discrete case.

$$\begin{aligned}
E[g(X)h(Y)] &= \sum_{x,\,y} g(x)h(y)p_{X,Y}(x,y) \\
&= \sum_x \sum_y g(x)h(y)p_X(x)p_Y(y) \\
&= \left[\sum_x g(x)p_X(x)\right]\left[\sum_y h(y)p_Y(y)\right] \\
&= E[g(X)]E[h(Y)].
\end{aligned}$$

7.2 Covariance, Variance of Sums, and Correlations

Theorem *If X and Y are independent, then* $\text{cov}(X,Y) = 0$.

Proof $\text{cov}(X,Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0.$

The reverse is not true.

Example A simple example of two dependent random variables X and Y having zero covariance is obtained by letting X be a random variable such that

$$P(X=0) = P(X=1) = P(X=-1) = \frac{1}{3}$$

and define

$$Y = \begin{cases} 0, & \text{if } X \neq 0 \\ 1, & \text{if } X = 0 \end{cases}.$$

Now, $XY = 0$, so $E[XY] = 0$. Also, $E[X] = 0$ and thus

$$\text{cov}(X,Y) = E[XY] - E[X]E[Y] = 0.$$

However, X and Y are clearly not independent.

7.2 Covariance, Variance of Sums, and Correlations

Properties

(i) $\mathrm{var}(X) = \mathrm{cov}(X, X)$.

(ii) $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$.

(iii) $\mathrm{cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \,\mathrm{cov}(X_i, Y_j)$.

Note: these applies for general situation – No independence is assumed.

Proof (i) This part is obvious. (ii) This is also obvious.

(iii) We first prove $\mathrm{cov}\left(\sum_{i=1}^{n} a_i X_i, Z\right) = \sum_{i=1}^{n} a_i \,\mathrm{cov}(X_i, Z)$. Note first that $E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \mu_{X_i}$.

Then
$$\begin{aligned}
\mathrm{cov}\left(\sum_{i=1}^{n} a_i X_i, Z\right) &= E\left[\left(\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i \mu_{X_i}\right)(Z - \mu_Z)\right] \\
&= E\left[\sum_{i=1}^{n} a_i (X_i - \mu_{X_i})(Z - \mu_Z)\right] = \sum_{i=1}^{n} a_i E[(X_i - \mu_{X_i})(Z - \mu_Z)] = \sum_{i=1}^{n} a_i \,\mathrm{cov}(X_i, Z).
\end{aligned}$$

The general case follows from
$$\mathrm{cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n} a_i \,\mathrm{cov}\left(X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n} a_i \sum_{j=1}^{m} b_j \,\mathrm{cov}(X_i, Y_j).$$

7.2 Covariance, Variance of Sums, and Correlations

Theorem

$$\operatorname{var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \operatorname{var}(X_k) + 2\sum_{1\le i<j\le n} \operatorname{cov}(X_i, X_j).$$

Proof

$$\begin{aligned}
\operatorname{var}\left(\sum_{k=1}^{n} X_k\right) &= \operatorname{cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \operatorname{cov}(X_i, X_j) \\
&= \sum_{i=1} \operatorname{cov}(X_i, X_i) + \sum_{1\le i\neq j\le n} \operatorname{cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} \operatorname{var}(X_i) + 2\sum_{1\le i<j\le n} \operatorname{cov}(X_i, X_j).
\end{aligned}$$

7.2 Covariance, Variance of Sums, and Correlations

Properties

(Variance of a Sum under Independence)
Let $X_1, \ldots, X_n$ be **independent** random variables, then

$$\operatorname{var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \operatorname{var}(X_k).$$

In other words, **under independence, variance of sum = sum of variances.**

7.2 Covariance, Variance of Sums, and Correlations

Example (Sample Variance)

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having expected value μ and variance σ^2, and let $\overline{X} = \sum_{i=1}^n X_i/n$ be the sample mean. The quantities $X_i - \overline{X}$, $i = 1, \ldots, n$, are called **deviations**, as they equal the differences between the individual data and the sample mean. The random variable

$$S^2 = \sum_{i=1}^n \frac{(X_i - \overline{X})^2}{n-1}$$

is called the **sample variance**. Find (a) $\mathrm{var}(\overline{X})$ and (b) $E[S^2]$.

Solution

(a)

$$\begin{aligned}
\mathrm{var}(\overline{X}) &= \left(\frac{1}{n}\right)^2 \mathrm{var}\left(\sum_{i=1}^n X_i\right) \\
&= \left(\frac{1}{n}\right)^2 \sum_{i=1}^n \mathrm{var}(X_i) \quad \text{by independence} \\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

7.2 Covariance, Variance of Sums, and Correlations

Solution (cont.)

(b) We start with the following algebraic identity

$$(n-1)S^2 = \sum_{i=1}^{n} (X_i - \mu + \mu - \overline{X})^2$$

Trick ☆

$$= \sum_{i=1}^{n} (X_i - \mu)^2 + \sum_{i=1}^{n} (\overline{X} - \mu)^2 - 2(\overline{X} - \mu) \sum_{i=1}^{n} (X_i - \mu)$$

$$= \sum_{i=1}^{n} (X_i - \mu)^2 + n(\overline{X} - \mu)^2 - 2(\overline{X} - \mu)n(\overline{X} - \mu) \qquad = \sum_{i=1}^{n} (X_i - \mu)^2 - n(\overline{X} - \mu)^2.$$

Taking expectations of the above yields that

$$\begin{aligned}(n-1)E[S^2] &= \sum_{i=1}^{n} E[(X_i - \mu)^2] - nE[(\overline{X} - \mu)^2] \\ &= n\sigma^2 - n\,\mathrm{var}(\overline{X}) \\ &= (n-1)\sigma^2\end{aligned}$$

where the final equality is due to part (a) and preceding one uses the the result that $E[\overline{X}] = \mu$. Dividing through by $n-1$ shows that

$$E[S^2] = \sigma^2.$$

7.2 Covariance, Variance of Sums, and Correlations

Remark This explains why

$$S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$$

is used as a estimator of σ^2 instead of the more "natural" choice of

$$\sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n}.$$

Example (Variance of Binomial)

Recall the representation of $X \sim \text{Bin}(n, p)$ given by $X = I_1 + I_2 + \cdots + I_n$. from section 7.1

We notice that the I_k's are independent. $I_k = \begin{cases} 1, & \text{if the kth trial is a success} \\ 0, & \text{if the kth trial is a failure} \end{cases}.$

Then,

$$\text{var}(X) = \text{var}\left(\sum_{k=1}^{n} I_k\right) = \sum_{k=1}^{n} \text{var}(I_k) = \sum_{k=1}^{n} p(1-p) = np(1-p).$$

Variance of bernoulli:

$X=1$: p success
$X=0$: $(1-p)$ failure.

$$E[X^2] - E[X]^2$$

$$= p - p^2 = p(1-p).$$

$E[X] = (1)p + 0(1-p) = p$

$E[X^2] = 1^2 p + 0^2(1-p) = p$

7.2 Covariance, Variance of Sums, and Correlations

Example (Variance of Number of Matches)

Recall the number of matches, X, is given as $X = \sum_{k=1}^{n} I_k$.

$$I_k = \begin{cases} 1, & \text{if the kth person gets back his hat} \\ 0, & \text{otherwise} \end{cases}.$$

Hence,

$$\begin{aligned} \operatorname{var}(X) &= \sum_{i=1}^{n} \operatorname{var}(I_i) + 2 \sum_{1 \le i < j \le n} \operatorname{cov}(I_i, I_j) \\ &= \sum_{i=1}^{n} \overset{p}{\frac{1}{n}} \overset{(1-p)}{\left(1 - \frac{1}{n}\right)} + 2 \sum_{1 \le i < j \le n} \operatorname{cov}(I_i, I_j) = n \cdot \frac{1}{n}(1 - 1/n) + n(n-1)\operatorname{cov}(I_1, I_2), \end{aligned}$$

$2 \times \binom{n}{2}$ 咁多 pair.

where

$$\begin{aligned} \operatorname{cov}(I_1, I_2) &= E(I_1 I_2) - E(I_1)E(I_2) = P(I_1 = I_2 = 1) - 1/n^2 \\ &= P(I_1 = 1)P(I_2 = 1 | I_1 = 1) - 1/n^2 = 1/n \times 1/(n-1) - 1/n^2 = \frac{1}{n^2(n-1)}. \end{aligned}$$

如1号已拿到，2号拿到的prob.

Therefore, $\operatorname{var}(X) = (1 - 1/n) + 1/n = 1.$

7.2 Covariance, Variance of Sums, and Correlations

Correlation and Correlation Coefficient

Definition *The **correlation (coefficient)** of random variables X and Y, denoted by $\rho(X,Y)$, is defined by*

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}}.$$

Theorem *We have,* $-1 \le \rho(X,Y) \le 1$.

Theorem $\mathrm{var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{var}(X_k) + 2 \sum_{1\le i<j\le n} \mathrm{cov}(X_i,X_j).$

Proof Suppose that X and Y have variances given by σ_X^2 and σ_Y^2, respectively. Then

$$0 \le \mathrm{var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = \frac{\mathrm{var}(X)}{\sigma_X^2} + \frac{\mathrm{var}(Y)}{\sigma_Y^2} + \frac{2\,\mathrm{cov}(X,Y)}{\sigma_X\sigma_Y} = 2[1+\rho(X,Y)]$$

implying that $-1 \le \rho(X,Y)$. On the other hand,

$$0 \le \mathrm{var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = \frac{\mathrm{var}(X)}{\sigma_X^2} + \frac{\mathrm{var}(Y)}{(-\sigma_Y)^2} - \frac{2\,\mathrm{cov}(X,Y)}{\sigma_X\sigma_Y} = 2[1-\rho(X,Y)]$$

implying that $\rho(X,Y) \le 1$, which completes the proof.

7.2 Covariance, Variance of Sums, and Correlations

Correlation and Correlation Coefficient

Remark (1) The correlation coefficient is a measure of the degree of linearity between X and Y. A value of $\rho(X,Y)$ near $+1$ or -1 indicates a high degree of linearity between X and Y, whereas a value near 0 indicates a lack of such linearity. A positive value of $\rho(X,Y)$ indicates that Y tends to increase when X does, whereas a negative value indicates that Y tends to decrease when X increases. If $\rho(X,Y) = 0$, then X and Y are said to be **uncorrelated**.

(2) $\rho(X,Y) = 1$ if and only if $Y = aX + b$ where $a = \sigma_Y/\sigma_X > 0$. ***why?***

(3) $\rho(X,Y) = -1$ if and only if $Y = aX + b$ where $a = -\sigma_Y/\sigma_X < 0$. ***why?***

(4) $\rho(X,Y)$ is dimensionless. *(meaning no measure unit like meter, cm, hour, minute, etc.)*

(5) Similar to covariance, if X and Y are independent, then $\rho(X,Y) = 0$.

Note that the converse is not true. In other words: If $\rho(X,Y) = 0$, meaning $\mathrm{cov}(X,Y) = 0$, then X and Y **may not be independent.**

7.2 Covariance, Variance of Sums, and Correlations

Correlation and Correlation Coefficient

Example Let I_A and I_B be indicator variables for the events A and B. That is,

$$I_A = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{otherwise} \end{cases},$$

$$I_B = \begin{cases} 1, & \text{if } B \text{ occurs} \\ 0, & \text{otherwise} \end{cases}.$$

Then

$$E[I_A] = P(A),$$

$$E[I_B] = P(B),$$

$$E[I_A I_B] = P(AB).$$

Therefore

$$\text{cov}(I_A, I_B) = P(AB) - P(A)P(B) = P(B)[P(A|B) - P(A)].$$

Thus we obtain the quite intuitive result that the indicator variables for A and B are either positively correlated, uncorrelated, or negatively correlated depending on whether $P(A|B)$ is, respectively, greater than, equal to, or less than $P(A)$.

7.2 Covariance, Variance of Sums, and Correlations

Correlation and Correlation Coefficient

Example Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having variance σ^2. Show that $\operatorname{cov}(X_i - \overline{X}, \overline{X}) = 0$.

Solution

We have

$$\operatorname{cov}(X_i - \overline{X}, \overline{X}) = \operatorname{cov}(X_i, \overline{X}) - \operatorname{cov}(\overline{X}, \overline{X})$$

$\bar{x}$ def.

$$= \operatorname{cov}\left(X_i, \frac{1}{n}\sum_{j=1}^{n} X_j\right) - \operatorname{var}(\overline{X})$$

抽出

$$= \frac{1}{n}\sum_{j=1}^{n} \operatorname{cov}(X_i, X_j) - \frac{\sigma^2}{n}$$

$\operatorname{cov}(X_i, X_j) = 0$ if $i \neq j$ (∵ independence)

$\operatorname{Cov}(X_i, X_j) = \operatorname{Var}(X_i)$ if $i = j$

$= \sigma^2$

$$= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0$$

where the next-to-last equality uses the result that $\operatorname{var}(\overline{X}) = \frac{\sigma^2}{n}$, and the final equality follows since

$$\operatorname{cov}(X_i, X_j) = \begin{cases} 0, & \text{if } j \neq i \text{ by independence} \\ \sigma^2, & \text{if } j = i \text{ since } \operatorname{var}(X_i) = \sigma^2 \end{cases}.$$

7.3 Conditional Expectation

Definition

(1) If X and Y are jointly distributed discrete random variables, then

$$E[X|Y=y] := \sum_x x p_{X|Y}(x|y), \qquad \text{if } p_Y(y) > 0.$$

(2) If X and Y are jointly distributed continuous random variables, then

$$E[X|Y=y] := \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, \mathrm{dx}, \qquad \text{if } f_Y(y) > 0.$$

7.3 Conditional Expectation

Example If X and Y are independent binomial random variables with identical parameters n and p, calculate the conditional expected value of X given that $X+Y=m$.

Solution Let us first calculate the conditional probability mass function of X, given that $X+Y=m$. For $k \le \min(n,m)$,

$$
\begin{aligned}
P(X=k|X+Y=m) &= \frac{P(X=k, X+Y=m)}{P(X+Y=m)} \\
&= \frac{P(X=k, Y=m-k)}{P(X+Y=m)} \\
&= \frac{P(X=k)P(Y=m-k)}{P(X+Y=m)} = \frac{\binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-m+k}}{\binom{2n}{m} p^m (1-p)^{2n-m}} = \frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}
\end{aligned}
$$

where we have used the fact that $X+Y$ is a binomial random variable with parameters $(2n, p)$. Hence the conditional distribution of X, given that $X+Y=m$, is the hypergeometric distribution; thus, we obtain

$$E[X|X+Y=m] = \frac{m}{2}.$$

$\frac{nm}{N}$, $N=2n$, $n=n$, $m=m$

7.3 Conditional Expectation

Example Suppose that X and Y have the joint probability density function

$$f(x,y) = \begin{cases} \frac{e^{-x/y}e^{-y}}{y}, & 0 < x < \infty,\ 0 < y < \infty \\ 0, & \text{otherwise} \end{cases}.$$

Compute $E[X|Y=y]$.

Solution It can be shown that

$$f_{X|Y}(x|y) = \frac{1}{y}e^{-x/y}.$$

Hence,

$$E[X|Y=y] = \int_0^\infty x f_{X|Y}(x|y)\,\mathrm{dx} = \int_0^\infty \frac{x}{y}e^{-x/y}\,\mathrm{dx} = y.$$

7.3 Conditional Expectation

Some Important Formulas

For any function g(x),

$$E[g(X)|Y=y] = \begin{cases} \sum_x g(x) p_{X|Y}(x|y), & \text{for discrete case} \\ \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\,\mathrm{dx}, & \text{for continuous case} \end{cases},$$

and hence

$$E\left[\sum_{k=1}^{n} X_k | Y=y\right] = \sum_{k=1}^{n} E[X_k|Y=y].$$

7.3.1 Computing Expectation by Conditioning

Notation: Let us denote by $E[X|Y]$ that function of the random variable Y whose value at $Y = y$ is $E[X|Y = y]$.

g(Y)=E[X|Y]

Example (1) Suppose that $E[X|Y = y] = y/2 - 5$, then $E[X|Y] = Y/2 - 5$.

(2) Suppose that $E[X|Y = y] = e^{y/2} - 5y$, then $E[X|Y] = e^{Y/2} - 5Y$.

Theorem

$$E[X] = E[E(X|Y)] = \begin{cases} \sum_y E(X|Y = y)P(Y = y), & \textit{if } Y \textit{ is discrete} \\ \int_{-\infty}^{\infty} E(X|Y = y) f_Y(y)\,\mathrm{d}y, & \textit{if } Y \textit{ is continuous} \end{cases}.$$

Proof We prove the result when X and Y are both continuous. Write $g(y) = E(X|Y = y)$, then

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

$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} x f_{X|Y}(x|y)\ \mathrm{d}x\right) f_Y(y)\ \mathrm{d}y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y)\ \mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\ \mathrm{d}x\,\mathrm{d}y$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\ \mathrm{d}y\,\mathrm{d}x = \int_{-\infty}^{\infty} x f_X(x)\ \mathrm{d}x = E[X].$$

7.3.1 Computing Expectation by Conditioning

Example A miner is trapped in a mine containing 3 doors. The first door leads to a tunnel that will take him to safety after 3 hours of travel. The second door leads to a tunnel that will return him to the mine after 5 hours of travel. The third door leads to a tunnel that will return him to the mine after 7 hours of travel. If we assume that the miner, on each choice (he "forgets his previous choices"), is equally likely to choose any one of the three doors, what is his expected length of time until he reaches for safety?

E[X]是从开始計算的 expected time

Solution Let X denote amount of time until he reaches for safety; Y denote the door he initially chooses.

Now
$$\begin{aligned} E[X] &= E[X|Y=1]P(Y=1)+E[X|Y=2]P(Y=2)+E[X|Y=3]P(Y=3) \\ &= \frac{1}{3}\left(E[X|Y=1]+E[X|Y=2]+E[X|Y=3]\right). \end{aligned}$$

Now,
$$\begin{aligned} E[X|Y=1] &= 3, \\ E[X|Y=2] &= 5+E[X], \\ E[X|Y=3] &= 7+E[X], \end{aligned}$$

在回到原點，再出發的 expected length 是 E[X].

hence $E[X]=\frac{1}{3}(3+5+E[X]+7+E[X])$. Solving this equation, we obtain $E[X]=15$.

7.3.1 Computing Expectation by Conditioning

Example (Expectation of a Random Sum).
Suppose $X_1, X_2, \ldots$ are independent and identically distributed with common mean μ. Suppose that N is a nonnegative integer valued random variable, independent of the X_k's. We are interested to find the mean of

$$T = \sum_{k=1}^{N} X_k, \qquad \text{where } T \text{ is taken to be zero when } N = 0.$$

Interpretations of the problem:

(a) N denotes the number of customers entering a department store during a period of time; X_k's amount spent by the kth customer; T total revenue.

(b) N denotes the number of claims against an insurance company; X_k's size of the kth claim; T total pay off.

Solution

$$\begin{aligned} E[T] &= E\left[\sum_{k=1}^{N} X_k\right] = \sum_{n=0}^{\infty} E\left[\sum_{k=1}^{N} X_k | N = n\right] P(N = n) \\ &= \sum_{n=1}^{\infty} E\left[\sum_{k=1}^{n} X_k | N = n\right] P(N = n) = \sum_{n=1}^{\infty} \left[\sum_{k=1}^{n} E(X_k)\right] P(N = n) \\ &= \sum_{n=1}^{\infty} n\mu P(N = n) = \mu E[N]. \end{aligned}$$

$E[T] = E(E(T|N))$ ← cond. prob. formula.

$= \sum_{n=0}^{\infty} E(T|N=n)\,P(N=n)$ ← discrete expectation formula

$= \sum_{n=0}^{\infty} E(X_1+X_2+\cdots+X_n)\,P(N=n)$ ← definition of T

$= \sum_{n=0}^{\infty} n\mu\,P(N=n)$ ← n個 r.v. 加總的 expectation 等於 n個 expection of r.v. (μ).

$= \mu E[N]$ ← 動用 $\sum_{n=0}^{\infty} nP(N=n) = E[N]$ formula.

7.3.1 Computing Expectation by Conditioning

Example Suppose that the number of people entering a department store on a given day is a random variable with mean 50. Suppose further that the amounts of money spent by those customers are independent random variables having a common mean of 8. Finally, suppose also that the amount of money spent by a customer is also independent of the total number of customers who enter the store. What is the expected amount of money spent in the store on a given day?

Solution

Using the previous example, the expected amount of money spent in the store is $50 \times \$8 = \$400$.

7.3.1 Computing Expectation by Conditioning

$N \sim Geo(p)$ $\quad P(N=n) = (1-p)^{n-1}p$

Example (Mean and variance of Geometric).

Let N be a geometrically distributed random variable with probability of success p. Let X_1 be the outcome of first trial.

$E[N \mid X_1 = \text{failure}]\,P(X_1 = \text{failure}) + E[N \mid X_1 = \text{success}]\,P(X_1 = \text{success})$

$$\begin{aligned} E[N] &= E[N|X_1=0]P(X_1=0) + E[N|X_1=1]P(X_1=1) \\ &= (1+E[N])q + 1 \times p. \end{aligned}$$

已知失败左一次，估计期望多久能成功.
答案=真期望值加一次

Solving for $E[N]$, we have $\quad E[N] = \dfrac{1}{p}$.

Similarly,

$$\begin{aligned} E[N^2] &= E[N^2|X_1=0]P(X_1=0) + E[N^2|X_1=1]P(X_1=1) \\ &= E[(1+N)^2]q + 1 \times p \\ &= 1 + 2qE[N] + qE[N^2] \\ &= 1 + \frac{2q}{p} + qE[N^2]. \end{aligned}$$

$E[N^2|X_1=0]$：已知fail左一次，expect是多少.
那麽代 $N+1$ 入 $E[f(n)]$ 中，即
$E[(N+1)^2]$.

Solving for $E[N^2]$, we get $\quad E[N^2] = \dfrac{1}{p} + \dfrac{2q}{p^2} = \dfrac{1+q}{p^2}$.

And hence, $\quad \mathrm{var}(N) = E[N^2] - (EN)^2 = \dfrac{q}{p^2}$.

7.3.2 Computing Probabilities by Conditioning

Not only can we obtain expectations by first conditioning on an appropriate random variable, but we may also use this approach to compute probabilities. To see this, let $X = I_A$ where A is an event. Then, we have

$$E(I_A) = P(A), \quad \text{and} \quad E(I_A|Y = y) = P(A|Y = y),$$

and by Theorem in Section 7.3.1.,

$$\begin{aligned} P(A) &= E(I_A) = E[E(I_A|Y)] \\ &= \begin{cases} \sum_y E(I_A|Y = y)P(Y = y), & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} E(I_A|Y = y) f_Y(y)\,\mathrm{d}y, & \text{if } Y \text{ is continuous} \end{cases} \end{aligned}$$

$$P(A) = \begin{cases} \sum_y P(A|Y = y)P(Y = y), & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} P(A|Y = y) f_Y(y)\,\mathrm{d}y, & \text{if } Y \text{ is continuous} \end{cases}.$$

This can be compared with the rule of total probabilities.

7.3.2 Computing Probabilities by Conditioning

Example Suppose that X and Y are independent continuous random variables having probability density functions f_X and f_Y respectively. Compute $P(X < Y)$.

Solution

Conditioning on the value of Y yields

$$\begin{aligned} P(X < Y) &= \int_{-\infty}^{\infty} P(X < Y | Y = y) f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} P(X < y | Y = y) f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} P(X < y) f_Y(y)\,\mathrm{d}y \quad \text{by independence} \\ &= \int_{-\infty}^{\infty} F_X(y) f_Y(y)\,\mathrm{d}y\,. \end{aligned}$$

7.3.2 Computing Probabilities by Conditioning

Example Suppose that X and Y are independent continuous random variables having probability density functions f_X and f_Y respectively. Find the distribution function of $X+Y$.

Solution By conditioning on the value of Y yields

$$\begin{aligned} P(X+Y\le a) &= \int_{-\infty}^{\infty} P(X+Y\le a|Y=y)f_Y(y)\,\mathrm{dy} \\ &= \int_{-\infty}^{\infty} P(X+y\le a|Y=y)f_Y(y)\,\mathrm{dy} \\ &= \int_{-\infty}^{\infty} P(X\le a-y)f_Y(y)\,\mathrm{dy} \qquad \text{by independence} \\ &= \int_{-\infty}^{\infty} F_X(a-y)f_Y(y)\,\mathrm{dy}\,. \end{aligned}$$

7.3.2 Computing Probabilities by Conditioning

Example Let U be a uniform random variable on (0,1), and suppose that the conditional density function of X, given that $U = p$, is binomial with parameters n and p. Find the probability mass function of X.

Solution

First note that X takes values $0, 1, \ldots, n$. For each $0 \leq k \leq n$, by conditioning on the value of U,

$$\begin{aligned} P(X=k) &= \int_0^1 P(X=k|U=p) f_U(p) \mathrm{dp} \\ &= \int_0^1 P(X=k|U=p) \mathrm{dp} \\ &= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} \mathrm{dp} \\ &= \binom{n}{k} \frac{k!(n-k)!}{(n+1)!} \quad \text{(claim below).} \end{aligned}$$

$$= \frac{n!}{k!\,(n-k)!} \cdot \frac{k!\,(n-k)!}{(n+1)!}$$

Claim. $B(k, n-k) := \int_0^1 p^k (1-p)^{n-k} \mathrm{dp} = \frac{k!(n-k)!}{(n+1)!}.$

next page

$$\int_0^1 \underset{v'}{p^k}\ \underset{u}{(1-p)^{n-k}}\,dp,$$

$u' = -(n-k)(1-p)^{n-k-1}$

$v = \dfrac{p^{k+1}}{k+1}$

$$\left[\frac{(1-p)^{n-k}p^{k+1}}{k+1}\right]_0^1 + \frac{n-k}{k+1}\int_0^1 (1-p)^{n-k-1}p^{k+1}\,dp$$

$$= \frac{n-k}{k+1}\beta(k+1, n-k-1)$$

再拆成: $\dfrac{n-k}{k+1}\ \dfrac{n-k-1}{k+2}$ 再变 factorial.

7.3.2 Computing Probabilities by Conditioning

Solution (cont.)

Proof of claim:

$$\begin{aligned} B(k,n-k) &= \int_0^1 (1-p)^{n-k} \mathrm{d}[p^{k+1}/(k+1)] \\ &= \frac{(n-k)}{(k+1)} \int_0^1 p^{k+1}(1-p)^{n-k-1} \mathrm{dp} \\ &= \frac{(n-k)}{(k+1)} B(k+1, n-k-1) \\ &= \frac{(n-k)(n-k-1)}{(k+1)(k+2)} B(k+2, n-k-2) \quad \cdots \quad = \frac{(n-k)!}{(k+1)(k+2)\cdots n} B(n,0) = \frac{k!(n-k)!}{(n+1)!} \end{aligned}$$

as $B(n,0) = \int_0^1 p^n dp = 1/(n+1)$.

So we obtain

$$P(X=k) = \frac{1}{n+1}, \quad k = 0, 1, \ldots, n.$$

That is, if a coin whose probability of coming up head is uniformly distributed over $(0,1)$ is flipped n times, then the number of heads occurring is equally likely to be any of the values $0, 1, \ldots, n$.

7.4. Conditional Variance

Definition *The* ***conditional variance*** *of X given that $Y = y$ is defined as*

$$\mathrm{var}(X|Y) \equiv E[(X - E[X|Y])^2|Y]$$

A very useful relationship between $\mathrm{var}(X)$ and $\mathrm{var}(X|Y)$ is given by

Theorem

$$\mathrm{var}(X) = E[\mathrm{var}(X|Y)] + \mathrm{var}(E[X|Y])$$

Proof Note that

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

Taking expectation on both sides yields

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

Since $E[E[X|Y]] = E[X]$, we have

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

Summing the two expressions above yields the required result.

7.4. Conditional Variance

Example Suppose that by any time t the number of people that have arrived at a train depot is a Poisson random variable with mean λt. If the initial train arrives at the depot at a time (independent of when the passengers arrive) that is uniformly distributed over $(0, T)$, what are the mean and variance of the number of passengers who enter the train?

0到T車都有可能來!

指的是人，不是車：實際到的人數

0 $N(Y)$ Y T

r.v of 車到

Solution

For $t \geq 0$, let $N(t)$ denote the number of arrivals during the interval $(0, t)$, and Y the time at which the train arrives.

We are interested to compute the mean and variance of the random variable $N(Y)$. We condition on Y to get

$$\begin{aligned} E[N(Y)|Y=t] = E[N(t)|Y=t] &= E[N(t)] && \text{by the independence of } Y \text{ and } N(t) \\ &= \lambda t && \text{since } N(t) \text{ is Poisson}(\lambda t). \end{aligned}$$

Hence, $E[N(Y)|Y] = \lambda Y$. Taking expectations gives $E[N(Y)] = E[E[N(Y)|Y]] = \lambda E[Y] = \dfrac{\lambda T}{2}$ since $Y \sim U(0, T)$.

To obtain $\text{var}(N(Y))$, note that

$$\begin{aligned} \text{var}(N(Y)|Y=t) &= \text{var}(N(t)|Y=t) \\ &= \text{var}(N(t)) && \text{by independence} \\ &= \lambda t \end{aligned}$$

next page →

7.4. Conditional Variance

Solution (cont.)

So

$$\mathrm{var}(N(Y)|Y) = \lambda Y.$$

Using the conditional variance formula, and the fact that $E[N(Y)|Y] = \lambda Y$ and $\mathrm{var}(N(Y)|Y) = \lambda Y$, we obtain

$$\begin{aligned}\mathrm{var}(N(Y)) &= E[\mathrm{var}(N(Y)|Y)] + \mathrm{var}(E[N(Y)|Y]) \\ &= E[\lambda Y] + \mathrm{var}(\lambda Y) \\ &= \lambda \frac{T}{2} + \lambda^2 \frac{T^2}{12}\end{aligned}$$

where we have used the fact that $\mathrm{var}(Y) = T^2/12$.

7.5. Moment Generating Function

$E X^k$ = moment.

Definition *The **moment generating function** of random variable X, denoted by M_X, is defined as*

exponential moment.

$$M_X(t) = E\left[e^{tX}\right] = \begin{cases} \sum_x e^{tx} p_X(x), & \text{if } X \text{ is discrete with probability mass function } p_X \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)\,\mathrm{d}x, & \text{if } X \text{ is continuous with probability density function } f_X \end{cases}.$$

Why do we call such a function a moment generating function?

Theorem Because this function generates all the moments of this random variable X. Indeed, for $n \geq 0$,

$$E(X^n) = M_X^{(n)}(0) \quad \text{where} \quad M_X^{(n)}(0) := \frac{d^n}{\mathrm{d}t^n} M_X(t)\,|_{t=0}\,.$$

Proof By Taylor series expansion, we have

$$E\left[e^{tX}\right] = E\left[\sum_{k=0}^{\infty} \frac{(tX)^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{E(tX)^k}{k!} = \sum_{k=0}^{\infty} \frac{E(X^k)}{k!} t^k$$

k'th derivative

and $M_X(t) = \sum_{k=0}^{\infty} \frac{M_X^{(k)}(0)}{k!} t^k$. Equating coefficient of t^n, we get $M_X^{(n)}(0) = E(X^n)$.

by taylor expansion.

7.5. Moment Generating Function

Theorem (Multiplicative Property).

If X and Y are independent, then $M_{X+Y}(t) = M_X(t)M_Y(t)$.

Proof $E\left[e^{t(X+Y)}\right] = E\left[e^{tX}e^{tY}\right] = E\left[e^{tX}\right]E\left[e^{tY}\right] = M_X(t)M_Y(t)$.

Theorem (Uniqueness Property).

Let X and Y be random variables with their moment generating functions M_X and M_Y respectively. Suppose that there exists an $h > 0$ such that

$$M_X(t) = M_Y(t), \quad \forall t \in (-h, h),$$

then X and Y have the same distribution (i.e., $F_X = F_Y$; or $f_X = f_Y$.)

Proof is skipped.

7.5. Moment Generating Function

Example When $X \sim \text{Be}(p)$, $M(t) = 1 - p + pe^t$.

Solution $$E\left[e^{tX}\right] = e^{t\cdot 0}P(X=0) + e^{t\cdot 1}P(X=1) = (1-p) + pe^t.$$

Example When $X \sim \text{Bin}(n,p)$, $M(t) = (1-p+pe^t)^n$.

Use the representation formula of Binomial random variable (sum of i.i.d. Bernoulli)

$$E\left[e^{tX}\right] = E\left[e^{t(I_1+\cdots+I_n)}\right] = E\left[e^{tI_1}\cdots e^{tI_n}\right] = E\left[e^{tI_1}\right]\cdots E\left[e^{tI_n}\right] = (1-p+pe^t)^n.$$

Example When $X \sim \text{Geom}(p)$, $M(t) = \frac{pe^t}{1-(1-p)e^t}$.

Example When $X \sim \text{Poisson}(\lambda)$, $M(t) = \exp(\lambda(e^t - 1))$.

Solution $$E\left[e^{tX}\right] = \sum_{k=0}^{\infty} \frac{e^{tk}e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^t)^k}{k!} = e^{-\lambda}e^{\lambda e^t} = \exp\left(\lambda\left(e^t - 1\right)\right).$$

7.5. Moment Generating Function

Example When $X \sim U(\alpha,\beta)$, $M(t) = \frac{e^{\beta t}-e^{\alpha t}}{(\beta-\alpha)t}$.

Example When $X \sim \mathrm{Exp}(\lambda)$, $M(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$.

Solution $E\left[e^{tX}\right] = \int_0^\infty e^{tx}\lambda e^{-\lambda x}\,\mathrm{dx} = \lambda \int_0^\infty e^{-(\lambda - t)x}\,\mathrm{dx} = \frac{\lambda}{\lambda - t}$, for $t < \lambda$.

Note that $M(t)$ is only defined for $t < \lambda$.

Example When $X \sim N(\mu, \sigma^2)$, $M(t) = \exp(\mu t + \sigma^2 t^2/2)$.

Solution Suppose that Y is standard normal.

$$E\left[e^{tY}\right] = \int_{-\infty}^{\infty} e^{ty}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\,\mathrm{dy} = e^{t^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-(y-t)^2/2}\,\mathrm{dy} = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-[y^2-2ty]/2}\,\mathrm{dy} = e^{t^2/2}.$$

For $X \sim N(\mu, \sigma^2)$, we write $X = \sigma Y + \mu$ where $Y \sim N(0,1)$.

$$E\left[e^{tX}\right] = E\left[e^{t\mu + t\sigma Y}\right] = e^{\mu t}E\left|e^{(\sigma t)Y}\right| = e^{\mu t + \sigma^2 t^2/2}.$$

7.5. Moment Generating Function

Example Suppose that $M_X(t) = e^{3(e^t-1)}$. Find $P(X=0)$.

Solution The given moment generating function is from that of a Poisson and the parameter of the Poisson is λ. Hence, $X \sim \text{Poisson}(\lambda)$. Therefore,

$$P(X=0) = e^{-3}.$$

Example Sum of independent Binomial random variables with the same success probability is Binomial.

Solution Let $X \sim \text{Bin}(n,p)$; $Y \sim \text{Bin}(m,p)$; and X and Y are independent.

$$\begin{aligned} E\left[e^{t(X+Y)}\right] &= E\left[e^{tX}\right]E\left[e^{tY}\right] \\ &= \left[1-p+pe^t\right]^n\left[1-p+pe^t\right]^m \\ &= \left[1-p+pe^t\right]^{n+m} \end{aligned}$$

which is the moment generating function of a $\text{Bin}(n+m,p)$. Hence, by uniqueness theorem,

$$X+Y \sim \text{Bin}(n+m,p).$$

7.5. Moment Generating Function

Example Sum of independent Poisson random variables is Poisson.

Solution Proceed as in the binomial case.

Example Sum of independent normal random variables is normal.

Solution Let $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$. We assume that X and Y are independent.

$$\begin{aligned} E\left[e^{t(X+Y)}\right] = E\left[e^{tX}\right] E\left[e^{tY}\right] &= \exp(\mu_1 t + \sigma_1^2 t^2/2)\, \exp(\mu_2 t + \sigma_2^2 t^2/2) \\ &= \exp([\mu_1 + \mu_2]t + [\sigma_1^2 + \sigma_2^2]t^2/2), \end{aligned}$$

which is the moment generating function of a normal distribution with mean $\mu_1 + \mu_2$ and $\sigma_1^2 + \sigma_2^2$.

Example Find all the moments of the exponential distribution of parameter $\lambda > 0$.

Solution Let $X \sim \text{Exp}(\lambda)$. Recall that

$$E\left[e^{tX}\right] = \frac{\lambda}{\lambda - t} = \frac{1}{1 - t/\lambda} = \sum_{k=0}^{\infty} \frac{t^k}{\lambda^k} = \sum_{k=0}^{\infty} \frac{k!/\lambda^k}{k!} t^k.$$

Hence, $E\left[X^k\right] = \dfrac{k!}{\lambda^k}.$

$$= \frac{1}{\lambda^n}\Gamma(n+1) = \frac{n!}{\lambda^n} \tag{1}$$

using the Gamma Function $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$ and that $\Gamma(n+1) = n!$ for any positive integer n.

Using the Moment Generating Function

The moment generating function (mgf) is

$$M_Y(t) = \mathbb{E}\left(e^{tY}\right) = \int_{-\infty}^{\infty} e^{ty} f(y)\,d y = \int_0^\infty e^{ty}\,\lambda\,e^{-\lambda y}\,d y$$
$$= \lambda \int_0^\infty e^{(t-\lambda)y}\,d y = \frac{\lambda}{t-\lambda}\left[e^{(t-\lambda)y}\right]_0^\infty = \frac{\lambda}{\lambda - t} \qquad \text{for } t < \lambda$$

Recalling the geometric series

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n \qquad \text{for } |z| < 1$$

we can expand the mgf for $t < \lambda$ as

$$M_Y(t) = \frac{\lambda}{\lambda - t} = \frac{1}{1-\frac{t}{\lambda}} = \sum_{n=0}^{\infty}\frac{1}{\lambda^n}t^n \tag{2}$$

Observing that

$$M_Y(t) = \sum_{n=0}^{\infty}\frac{M_Y^{(n)}(0)}{n!}t^n = \sum_{n=0}^{\infty}\frac{\mathbb{E}(Y^n)}{n!}t^n \tag{3}$$

equating the coefficients of (2) and (3) we find

$$M_Y^{(n)}(0) = \mathbb{E}(Y^n) = \frac{n!}{\lambda^n}. \tag{4}$$

Solution

7.5. Moment Generating Function

Example A chi-squared random variable with n degrees of freedom, χ_n^2, is given as

$$Z_1^2 + \cdots + Z_n^2$$

where $Z_1, \ldots, Z_n$ are independent standard normal random variables. Compute the moment generating function of a chi-squared random variable with n degrees of freedom.

Solution Let $M(t)$ be its moment generating function. By the above form,

$$M(t) = \left(E\left[e^{tZ^2}\right]\right)^n \qquad \text{where } Z \text{ is a standard normal.}$$

Now,

$$\begin{aligned} E\left[e^{tZ^2}\right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx^2} e^{-x^2/2}\,dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2}\,dx, \quad \text{where } \sigma^2 = (1-2t)^{-1}. \\ &= \sigma \quad = (1-2t)^{-1/2}, \end{aligned}$$

where the next-to-last equality uses the fact that the normal density with mean 0 and variance σ^2 integrates to 1.

Therefore, $M(t) = (1-2t)^{-n/2}$.

7.6. Joint Moment Generating Functions

It is also possible to define the joint moment generating function of two or more random variables. This is done as follows. For any n random variables $X_1, \ldots, X_n$, the joint moment generating function, $M(t_1, \ldots, t_n)$, is defined for all real values of $t_1, \ldots, t_n$ by

$$M(t_1, \ldots, t_n) = E\left[e^{t_1 X_1 + \cdots + t_n X_n}\right].$$

The individual moment generating functions can be obtained from $M(t_1, \ldots, t_n)$ by letting all but one of the t_j's be 0. That is,

$$M_{X_i}(t) = E\left[e^{tX_i}\right] = M(0, \ldots, 0, t, 0, \ldots, 0),$$

where the t is in the ith place.

7.6. Joint Moment Generating Functions

It can be proved (although the proof is too advanced for this course) that $M(t_1, \ldots, t_n)$ uniquely determines the joint distribution of $X_1, \ldots, X_n$. This result can then be used to prove that the n random variables $X_1, \ldots, X_n$ are independent if and only if

$$M(t_1, \ldots, t_n) = M_{X_1}(t_1) \cdots M_{X_n}(t_n). \tag{7.5}$$

This follows because, if the n random variables are independent, then

$$\begin{aligned} M(t_1, \ldots, t_n) &= E\left[e^{t_1 X_1 + \cdots + t_n X_n}\right] \\ &= E\left[e^{t_1 X_1} \cdots e^{t_n X_n}\right] \\ &= E\left[e^{t_1 X_1}\right] \cdots E\left[e^{t_n X_n}\right] \quad \text{by independence} \\ &= M_{X_1}(t_1) \cdots M_{X_n}(t_n). \end{aligned}$$

On the other hand, if Equation (7.5) is satisfied, then the joint moment generating function $M(t_1, \ldots, t_n)$ is the same as the joint moment generating function of n independent random variables, the ith of which has the same distribution as X_i. As the joint moment generating function uniquely determines the joint distribution, this must be the joint distribution; hence the random variables are independent.

7.6. Joint Moment Generating Functions

Example Let X and Y be independent normal random variables, each with mean μ and variance σ^2. In Chapter 6, we showed that $X+Y$ and $X-Y$ are independent. Let us now establish this result by computing their joint moment generating function.

Solution

$$\begin{aligned}
E\left[e^{t(X+Y)+s(X-Y)}\right] &= E\left[e^{(t+s)X+(t-s)Y}\right] \\
&= E\left[e^{(t+s)X}\right] E\left[e^{(t-s)Y}\right] \\
&= e^{\mu(t+s)+\sigma^2(t+s)^2/2} e^{\mu(t-s)+\sigma^2(t-s)^2/2} \\
&= e^{2\mu t+\sigma^2 t^2} e^{\sigma^2 s^2}.
\end{aligned}$$

But we recognize the preceding as the joint moment generating function of the sum of a normal random variable with mean 2μ and variance $2\sigma^2$ and an independent normal random variable with mean 0 and variance $2\sigma^2$. As the joint moment generating function uniquely determines the joint distribution, it thus follows that $X+Y$ and $X-Y$ are independent normal random variables.

7.6. Joint Moment Generating Functions

Example Suppose that the number of events that occur is a Poisson random variable with mean λ, and that each event is independently counted with probability p. Show that the number of counted events and the number of uncounted events are independent Poisson random variables with means λp and $\lambda(1-p)$ respectively.

Solution

Let X denote the total number of events, and let X_c denote the number of them that are counted. To compute the joint moment generating function of X_c, the number of events that are counted, and $X - X_c$, the number that are uncounted, start by conditioning on X to obtain

$$E\left[e^{sX_c+t(X-X_c)} \mid X=n\right] = e^{tn}E\left[e^{(s-t)X_c} \mid X=n\right]$$
$$= e^{tn}\left(pe^{s-t}+1-p\right)^n = \left(pe^s+(1-p)e^t\right)^n$$

where the preceding equation follows since conditional on $X=n$, X_c is a binomial random variable with parameters n and p. Hence

next page

$$E\left[e^{sX_c+t(X-X_c)} \mid X\right] = \left(pe^s+(1-p)e^t\right)^X.$$

7.6. Joint Moment Generating Functions

Solution (cont.)

Taking expectations of both sides of the preceding yields that

$$E\left[e^{sX_c+t(X-X_c)}\right] = E\left[\left(pe^s + (1-p)e^t\right)^X\right].$$

Now, since X is Poisson with mean λ, it follows that $E\left[e^{tX}\right] = e^{\lambda(e^t-1)}$. Therefore, for any positive value a we see (by letting $a = e^t$) that $E\left[a^X\right] = e^{\lambda(a-1)}$. Thus

$$\begin{aligned} E\left[e^{sX_c+t(X-X_c)}\right] &= e^{\lambda(pe^s+(1-p)e^t-1)} \\ &= e^{\lambda p(e^s-1)}e^{\lambda(1-p)(e^t-1)}. \end{aligned}$$

As the preceding is the joint moment generating function of independent Poisson random variables with respective means λp and $\lambda(1-p)$, the result is proven.