Chapter 7 Properties of Expectation

7.1 Introduction

- **Discrete case**: $E[X] = \sum_x xp(x)$
- **Continuous case**: $E[X] = \int_{-\infty}^{\infty} xf(x)\,dx$
- If $P\{a \le X \le b\} = 1$, then

$$a \le E[X] \le b$$

7.2 Expectation of sums of random variables

Proposition 2.1: If X and Y have a joint probability mass function $p(x, y)$, then

$$E[g(X, Y)] = \sum_y \sum_x g(x, y)p(x, y)$$

If X and Y have a joint probability density function $f(x, y)$, then

$$E[g(x, y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)\,dxdy$$

Example 7.2a. 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.

- $f(x, y) = \frac{1}{L^2}, \quad 0 < x < L, \quad 0 < y < L$
- $E[|X - Y|] = \dfrac{1}{L^2} \int_0^L \int_0^L |x - y| \, dy dx$
- Now,

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

- Therefore,

$$\begin{aligned} E[|X-Y|] &= \frac{1}{L^2} \int_0^L \left(\frac{L^2}{2} + x^2 - xL \right) dx \\ &= \frac{L}{3} \end{aligned}$$

-
$$
\begin{aligned}
E[X+Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)\,dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,dydx + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\,dxdy \\
&= \int_{-\infty}^{\infty} xf_X(x)\,dx + \int_{-\infty}^{\infty} yf_Y(y)\,dy \\
&= E[X]+E[Y]
\end{aligned}
$$

- $E[X+Y]=E[X]+E[Y]$ if $E[X]$ and $E[Y]$ are finite.

Example 7.2b. Suppose that for random variables X and Y,

$$X \geq Y$$

That is, for any outcome of the probability experiment, the value of the random variable X is greater than or equal the value of the random variable Y. Since the preceding is equivalent to the inequality $X - Y \geq 0$, it follows that $E[X-Y] \geq 0$, or, equivalently,

$$E[X] \geq E[Y]$$

If $E[X_i]$ is finite for all $i = 1, \ldots, n$, then

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$$

Example 7.2c. *The 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. The quantity $\overline{X}$, defined by

$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

is called the **sample mean**. Compute $E[\overline{X}]$.

•

$$\begin{aligned} E[\overline{X}] &= E\left[\sum_{i=1}^{n} \frac{X_i}{n}\right] \\ &= \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n} \sum_{i=1}^{n} E[X_i] \\ &= \mu \qquad \text{since } E[X_i] \equiv \mu \end{aligned}$$

Example 7.2d. *Boole's inequality.* Let $A_1, \ldots, A_n$ denote events and defined the indicator variables X_i, $i = 1, \ldots, n$ by

$$X_i = \begin{cases} 1 & \text{if } A_i \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

- $X = \sum_{i=1}^{n} X_i$: The number of the events A_i that occur.
- Let

$$Y = \begin{cases} 1 & \text{if } X \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

- Y is equal to 1 if at least one of the A_i occurs and is 0 otherwise.
- Then $X \geq Y$ and $E[X] \geq E[Y]$.
- But since

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} P(A_i)$$

and

$$E[Y] = P\{\text{at least one of the } A_i \text{ occur}\} = P\left(\bigcup_{i=1}^{n} A_i\right)$$

- We obtain Boole's inequality

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

Next three examples show how Eq. (2.2) can be used to calculate the expected value of binomial, negative binomial, and hypergeometric random variables.

Example 7.2e. *Expectation of a binomial random variable.* Let X be a binomial random variable with parameters n and p.

- Note that $X = X_1 + X_2 + \cdots + X_n$ where
$$X_i = \begin{cases} 1 & \text{if the } i\text{th trial is a success} \\ 0 & \text{if the } i\text{th trial is a failure} \end{cases}$$
- Hence, X_i is a Bernoulli random variable having expectation $E[X_i] = 1(p) + 0(1 - p)$.
- Thus $E[X] = E[X_1] + \cdots + E[X_n] = np$

Example 7.2f. *Mean of a negative binomial random variable.* 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.

- X: The number of trials needed to amass a total of r successes.
- X_i: The number of additional trials required, after the $(i-1)$st success, until a total of i successes are amassed.
- Note that $X = X_1 + X_2 + \cdots + X_r$
- $E[X] = E[X_1] + \cdots + E[X_r] = \frac{r}{p}$

Example 7.2g. *Mean of a hypergeometric random variable.* 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.

- X: The number of white balls selected.
- $X = X_1 + \cdots + X_m$ where
$$X_i = \begin{cases} 1 & \text{if the } i\text{th white ball is selected} \\ 0 & \text{otherwise} \end{cases}$$
- Now,
$$E[X_i] = P\{X_i = 1\}$$

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

- Hence

$$E[X] = E[X_1] + \cdots + E[X_m] = \frac{mn}{N}$$

- Alternative: $X = Y_1 + \cdots + Y_n$ where

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

- $E[Y_i] = \frac{m}{N}$
- Then $E[X] = E[Y_1] + \cdots + E[Y_n] = \frac{nm}{N}$

Example 7.2h. *Expected number of matches.* A group of N people throw their hats into the center of a room. The hats are mixed up, and each person randomly selected one. Find the expected number of people that select their own hats.

- X: The number of matches.
- $X = X_1 + X_2 + \cdots + X_N$ where

  $$X_i = \begin{cases} 1 & \text{if the } i\text{th person selects his own hat} \\ 0 & \text{otherwise} \end{cases}$$

- $E[X_i] = P\{X_i = 1\} = \frac{1}{N}$
- Then $E[X] = E[X_1] + \cdots + E[X_N] = (\frac{1}{N}) = 1$

Example 7.2i. The following problem was posed and solved in the eighteenth century by Daniel Bernoulli. Suppose that a jar contains $2N$ cards, two of them marked 1, two marked 2, two marked 3, and so on. Draw out m cards at random. What is the expected number of pairs that still remain in the jar? (Interestingly enough, Bernoulli proposed the above as a possible probabilistic model for determining the number of marriages that remain intact when there is a total of m deaths among the N married couples.)

- Define for $i = 1, 2, \ldots, N$,

$$X_i = \begin{cases} 1 & \text{if the } i\text{th pair remains in the jar} \\ 0 & \text{otherwise} \end{cases}$$

- Now,

$$\begin{aligned} E[X_i] &= P\{X_i = 1\} \\ &= \frac{\binom{2N-2}{m}}{\binom{2N}{m}} \\ &= \frac{\frac{(2N-2)!}{m!(2N-2-m)!}}{\frac{(2N)!}{m!(2N-m)!}} \\ &= \frac{(2N-m)(2N-m-1)}{(2N)(2N-1)} \end{aligned}$$

- Hence the desired result is

$$\begin{aligned} E[X_1 + \cdots + X_N] &= E[X_1] + \cdots + E[X_N] \\ &= \frac{(2N-m)(2N-m-1)}{2(2N-1)} \end{aligned}$$

Example 7.2j. *Coupon-collecting problems.* 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.

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

- X: The number of different types of coupons in the set of n coupons.

(a) $X = X_1 + \cdots + X_N$ where

$$X_i = \begin{cases} 1 & \text{if at least one type } i \text{ coupon is contained in the set of } n \\ 0 & \text{otherwise} \end{cases}$$

- Now,

$$\begin{aligned} E[X_i] &= P\{X_i = 1\} \\ &= 1 - P\{\text{no type } i \text{ coupon are contained in the set of } n\} \\ &= 1 - \left(\frac{N-1}{N}\right)^n \end{aligned}$$

- Hence

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

(b) Y: The number of coupons collected before a complete set is attained.

- Y_i: The number of additional coupons that need to be obtained after i distinct types have been collected in order to obtain another distinct type.
- $Y = Y_0 + Y_1 + \cdots + Y_{N-1}$
- $P\{Y_i = k\} = \frac{N-i}{N}(\frac{i}{N})^{k-1} \quad k \geq 1$
- $Y_i \sim \text{geometric}((N-i)/N)$
- $E[Y_i] = \frac{N}{N-i}$
- Then

$$\begin{aligned} E[Y] &= 1 + \frac{N}{N-1} + \frac{N}{N-2} + \cdots + \frac{N}{1} \\ &= N\left[1 + \cdots + \frac{1}{N-1} + \frac{1}{N}\right] \\ &\approx N(\log N + C) \end{aligned}$$

where $C \approx 0.57721$ is the Euler constant.

Example 7.2k. Ten hunters are waiting for ducks to fly by. When a flock of ducks flies overhead, the hunters fire at the same time, but each chooses his target at random, independently of the others. If each hunter independently hits his target with probability p, compute the expected number of ducks that escape unhurt when a flick of size 10 flies overhead.

- X_i: 1 if the ith duck escapes unhurt and 0 otherwise.
- $E[X_1 + \cdots + X_{10}] = E[X_1] + \cdots + E[X_{10}]$
- Each of the hubters will hit the ith duck with probability $p/10$.
- $P\{X_i = 1\} = \left(1 - \frac{p}{10}\right)^{10}$
- $E[X] = 10\left(1 - \frac{p}{10}\right)^{10}$

Example 7.2l. *Expected number of runs.* Suppose that a sequence of n 1's and m 0's is randomly permuted so that each of the $(n+m)!/(n!m!)$ possible arrangements is equally likely. Any consecutive string of 1's is said to constitute a run of 1's-for instance, if $n = 6, m = 4$, and the ordering is 1,1,1,0,1,1,0,0,1,0, then there are 3 runs of 1's-and we are interested in computing the mean number of such runs.

- Let
$$I_i = \begin{cases} 1 & \text{if a run of 1's starts at the } i\text{th position} \\ 0 & \text{otherwise} \end{cases}$$
- $R(1)$: The number of runs of 1.
$$\begin{aligned} R(1) &= \sum_{i=1}^{n+m} I_i \\ E[R(1)] &= \sum_{i=1}^{n+m} E[I_i] \end{aligned}$$
- Now,
$$\begin{aligned} E[I_1] &= P\{\text{"1" in position } 1\} \\ &= \frac{n}{n+m} \end{aligned}$$
and for $1 < i \le n + m$,
$$\begin{aligned} E[I_i] &= P\{\text{"0" in position } i-1, \text{"1" in position } i\} \\ &= \frac{m}{n+m}\frac{n}{n+m-1} \end{aligned}$$

- Hence
$$E[R(1)] = \frac{n}{n+m} + \frac{(n+m-1)nm}{(n+m)(n+m-1)}$$
- Similarly, $E[R(0)]$, the expected number of runs of 0's, is
$$E[R(0)] = \frac{m}{n+m} + \frac{nm}{n+m}$$
and the expected number of runs of either type is
$$E[R(1) + R(0)] = 1 + \frac{2nm}{n+m}$$

Example 7.2m. Consider an ordinary deck of cards that is turned face up one card at a time. How many cards would one expect to turn face up in order to obtain (a) an ace and (b) a spade?

- (a) and (b) are special cases of the following problem.
- Suppose that balls are taken one by one out of an urn containing n white and m black balls until the first white ball is drawn.

- If X denotes the number of balls withdrawn.
- Name the black balls as $b_1, \ldots, b_m$.
- Let

$$X_i = \begin{cases} 1 & \text{if } b_i \text{ is withdrawn before any of the white balls} \\ 0 & \text{otherwise} \end{cases}$$

- $X = 1 + \sum_{i=1}^{m} X_i$
- Hence $E[X] = 1 + \sum_{i=1}^{m} P\{X_i = 1\}$
- As each of these n white balls plus ball b_i has an equal probability of being the first one of this set to be withdrawn

$$E[X_i] = P\{X_i = 1\} = \frac{1}{n+1}$$

- $E[X] = 1 + \frac{m}{n+1}$

Example 7.2n. *A random walk in the plane.* Consider a particle initially located at a given point in the plane and suppose that it undergoes a sequence of steps of fixed length but in a completely random direction. Specifically, suppose that the new position after each step is one unit of distance from the previous position and at an angle of orientation from the previous position that is uniformly distributed over $(0, 2\pi)$ (see Fig. 7.3). Compute the expected square of the distance from the origin after n steps.

- (X_i, Y_i): The change in the position at the ith step.
- $X_i = \cos\theta_i \quad Y_i = \sin\theta_i$
- $$\begin{aligned} D^2 &= \left(\sum_{i=1}^{n} X_i\right)^2 + \left(\sum_{i=1}^{n} Y_i\right)^2 \\ &= \sum_{i=1}^{n}(X_i^2 + Y_i^2) + \sum\sum_{i\neq j}(X_iX_j + Y_iY_j) \\ &= n + \sum\sum_{i\neq j}(\cos\theta_i\cos\theta_j + \sin\theta_i\sin\theta_j) \end{aligned}$$
- $$\begin{aligned} 2\pi E[\cos\theta_i] &= \int_0^{2\pi}\cos u\,du = \sin 2\pi - \sin 0 = 0 \\ 2\pi E[\sin\theta_i] &= \int_0^{2\pi}\sin u\,du = \cos 0 - \cos 2\pi = 0 \end{aligned}$$
- $E[D^2] = n$

Example 7.2o. *Analyzing the quick-sort algorithm.* Suppose that we are presented with a set of n distinct values $x_1, \ldots, x_n$ and we desire to put them in increasing order. An efficient procedure for accomplishing this task is the quick-sort algorithm.

- When $n = 2$, the algorithm compares the two values and then put them in the appropriate order.
- When $n > 2$, one of the elements is randomly chosen–say it is x_i–and then all of the other values are compared with x_i.
- the algorithm then repeats itself on these brackets and continues until all values have been sorted.
- Example: 5, 9, 3, 10, 11, 14, 8, 4, 17, 6
  - $\{5, 9, 3, 8, 4, 6\}, 10, \{11, 14, 17\}$
  - $\{5, 3, 4\}, 6, \{9, 8, \}, 10, \{11, 14, 17\}$
  - $\{3\}, 4, \{5\}, 6, \{9, 8, \}, 10, \{11, 14, 17\}$

– This continues until there is no bracketed set that contains more than a single value.

- X: The number of comparisons that it takes the quick-sort algorithm to sort n distinct numbers, then $E[X]$ is a measure of the effectiveness of this algorithm.
- $I(i,j)$: 1 if i and j are ever directly compared, 0 otherwise.
- $X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I(i,j)$
- $E[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P\{i \text{ and } j \text{ are ever compared}\}$
- $P\{i \text{ and } j \text{ are ever compared}\} = \dfrac{2}{j-i+1}$
- $E[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \dfrac{2}{j-i+1}$

$$\begin{aligned} \sum_{j=i+1}^{n} \frac{2}{j-i+1} &\approx \int_{i+1}^{n} \frac{2}{x-i+1}\,dx \\ &= 2\log(n-i+1) - 2\log(2) \\ &\approx 2\log(n-i+1) \end{aligned}$$

$$\begin{aligned} E[X] &\approx \sum_{i=1}^{n} 2\log(n-i+1) \\ &\approx 2\int_1^{n-1} \log(n-x+1)\,dx \\ &= 2\int_2^{n} \log(y)\,dy \\ &\approx 2n\log(n) \end{aligned}$$

Example 7.2p. *The probability of a union of events.* Let $A_1, \ldots, A_n$ denote events and

$$X_i = \begin{cases} 1 & \text{if } A_i \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

- $1 - \prod_{i=1}^{n}(1 - X_i) = 1$ if $\cup A_i$ occurs and 0 otherwise.
- $E\left[1 - \prod_{i=1}^{n}(1-X_i)\right] = P\left(\bigcup_{i=1}^{n} A_i\right)$
- $\prod_{i=1}^{n}(1-X_i) = \sum_{k=0}^{n}(-1)^k \sum_{i_1<\cdots<i_k} X_{i_1}\cdots X_{i_k}$
- $E[X_{i_1}X_{i_2}\cdots X_{i_k}] = P(A_{i_1}A_{i_2}\cdots A_{i_k})$
- $E\left[1 - \prod_{i=1}^{n}(1-X_i)\right] = \sum_{k=1}^{n}(-1)^{k+1}\sum_{i_1<\cdots<i_k} P(A_{i_1}\cdots A_{i_k})$

Example 7.2q. A round-robin tournament of n contestants is one in which each of the

$\binom{n}{2}$ pairs of contestants play each other exactly once, with the outcome of any play being that one of the contestants wins and the other loses.

- Suppose that the n players are initially numbered as player 1, player 2, and so on.
- Hamiltonian permutation if i_1 beats i_2, i_2 beats $i_3, \ldots,$ and i_{n-1} beats i_n.
- A problem of some interest is to determine the largest possible number of Hamiltonian permutations.
- Suppose that there are 3 players. Then it easy to see that if one of the players wins twice, then there is a single Hamiltonian permutation.
- If each of the players wins once, then there will be three Hamiltonians.
- We will introduce randomness to show that in a round-robin tournament of n players, $n > 2$, there is an outcome for which the

number of Hamiltonian permutations is greater than $n!/2^{n-1}$.

- Suppose that the results of the $\binom{n}{2}$ games are independent and that either of the two contestants is equally likely to win each encounter.
- X: The number of Hamiltonians that result.
- Since at least one of the possible values of a nonrandom variable must exceed its mean, it follows that there must be at least one possible tournament result which has more than $E[X]$ Halmiltonian permutations.
- To determine $E[X]$, number of the $n!$ permutations, for $i = 1, \ldots, n!$, $X_i = 1$ if permutation i is a Halmitonian, 0 otherwise.
- $E[X] = \sum_i E[X_i]$
- $E[X_i] = (1/2)^{n-1}$
- $E[X] = \frac{n!}{2^{n-1}}$

$$E\left[\lim_{n\to\infty}\sum_{i=1}^{n} X_i\right] = \lim_{n\to\infty} E\left[\sum_{i=1}^{n} X_i\right]$$

holds in two important special cases:

1. The X_i are all nonnegative random variables.
2. $\sum_{i=1}^{\infty} E[|X_i|] < \infty$

Example 7.2r. Consider any nonnegative, integer-valued random variable X.

- If for each $i \geq 1$, we define

$$X_i = \begin{cases} 1 \text{ if } X \geq i \\ 0 \text{ if } X < i \end{cases}$$

then

$$\begin{aligned} \sum_{i=1}^{\infty} X_i &= \sum_{i=1}^{X} X_i + \sum_{i=X+1}^{\infty} X_i \\ &= \sum_{i=1}^{X} 1 + \sum_{i=X+1}^{\infty} 0 \\ &= X \end{aligned}$$

- Hence, since the X_i are all nonnegative,

$$\begin{aligned} E[X] &= \sum_{i=1}^{\infty} E[X_i] \\ &= \sum_{i=1}^{\infty} P\{X \geq i\} \end{aligned}$$

  a useful identity.

Example 7.2s. Suppose that n elements-call them $1, 2, \ldots, n$-must be stored in a computer in the form of an ordered list. Each unit of time a request will be made for one of these elements-i being requested, independently of the past, with probability $P(i), i \geq 1, \sum_i P(i) = 1$. Assuming these probabilities are known, what ordering minimizes the average position on the line of the element requested?

- Suppose that the elements are numbered so that $P(1) \geq P(2) \geq \cdots \geq P(n)$.
- To show that $1, 2, \ldots, n$ is the optimal ordering, let X denote the position of the requested element.

- Now under any ordering say $O = i_1, i_2, \ldots, i_n$,

$$\begin{aligned} P_O\{X \geq k\} &= \sum_{j=k}^{n} P(i_j) \\ &\geq \sum_{j=k}^{n} P(j) \\ &= P_{1,2,\ldots,n}\{X \geq k\} \end{aligned}$$

- $E_O[X] \geq E_{1,2,\ldots,n}[X]$

7.3 Covariance, variance of sums, and correlations

Proposition 3.1: If X and Y are independent, then for any functions h and g,

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

Definition: The covariance between X and Y, denoted by $\mathrm{Cov}(X, Y)$, is defined by

$$\begin{aligned} \mathrm{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

A simple example of two dependent random variables X and Y have zero covariance:

- $P\{X = 0\} = P\{X = 1\} = P\{X = -1\} = \frac{1}{3}$
- $Y = 0$ if $X \neq 0$ and 1 if $X = 0$.
- $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0$ since $E[XY] = 0$ and $E[X] = 0$.

Proposition 3.2:

(i) $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$

(ii) $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$

(iii) $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$

(iv) $\mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} \mathrm{Cov}(X_i, Y_j)$

$$\boxed{\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{i<j} \mathrm{Cov}(X_i, X_j)}$$

If $X_1, \ldots, X_n$ are pairwise independent, then

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Example 7.3a. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having expected value μ and variance σ^2, and as in Example 2c, 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) $E[S^2]$

(a)

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

(b)

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

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

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

Example 7.3b. *Variance of a binomial random variable.* Compute the variance of a binomial random variable X with parameters n and p.

- $X = X_1 + \cdots + X_n$ where
$$X_i = \begin{cases} 1 & \text{if the } i\text{th trial is a success} \\ 0 & \text{otherwise} \end{cases}$$
- $\mathrm{Var}(X) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)$
- $$\mathrm{Var}(X_i) = E[X_i^2] - (E[X_i])^2$$

$$\begin{aligned} &= E[X_i] - (E[X_i])^2 \qquad \text{since } X_i^2 = X_i \\ &= p - p^2 \end{aligned}$$

- $\mathrm{Var}(X) = np(1-p)$

Example 7.3c. *Variance of the number of matches.* Compute the variance of X, the number of people that select their own hats in Example 2h.

- $X = X_1 + \cdots + X_N$ where

$$X_i = \begin{cases} 1 & \text{if the } i\text{th man selects his own hat} \\ 0 & \text{otherwise} \end{cases}$$

- $\mathrm{Var}(X) = \sum_{i=1}^{N} \mathrm{Var}(X_i) + 2 \sum \sum_{i<j} \mathrm{Cov}(X_i, X_j)$
- $\mathrm{Var}(X_i) = \frac{1}{N}(1 - \frac{1}{N}) = \frac{N-1}{N^2}$
- $\mathrm{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$

$$X_i X_j = \begin{cases} 1 & \text{if the } i\text{th and } j\text{th men both select their own hats} \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E[X_iX_j] &= P\{X_i = 1, X_j = 1\} \\ &= P\{X_i = 1\}P\{X_j = 1|X_i = 1\} \\ &= \frac{1}{N}\frac{1}{N-1} \end{aligned}$$

- $\text{Cov}(X_i, X_j) = \frac{1}{N(N-1)} - \left(\frac{1}{N}\right)^2 = \frac{1}{N^2(N-1)}$
-

$$\begin{aligned} \text{Var}(X) &= \frac{N-1}{N} + 2\binom{N}{2}\frac{1}{N^2(N-1)} \\ &= \frac{N-1}{N} + \frac{1}{N} \\ &= 1 \end{aligned}$$

Example 7.3d. *Sampling from a finite population.* Consider a set N people each of whom has an opinion about a certain subject that is measured by a real number v, which represents the person's "strength of feeling" about the subject. Let v_i represent the strength of feeling of person $i, i = 1, \ldots, N$. Suppose that these quantities $v_i, i = 1, \ldots, N$ are unknown and to gather information a group of n of the N people is *"randomly chosen"* in the sense that all of the $\binom{N}{n}$ subsets of size n are equally likely to be chosen. These n people are then questioned and their feelings determined. If S denotes the sum of the n sampled values, determine its mean and variance.

An important application of the above is to a forthcoming election in which each person in the population is either for or against a certain candidate or proposition. If we take v_i to equal 1 if person i is in favor and 0 if he or she is against, then $\overline{v} = \sum_{i=1}^{N} v_i/N$ represents the proportion of the population that is in favor. To estimate $\overline{v}$, a random sample of n people is chosen, and these people are polled. The proportion of those polled that are in favor-that is, S/n-is often used used as an estimate of $\overline{v}$.

- $$I_i = \begin{cases} 1 & \text{if person } i \text{ is in the random sample} \\ 0 & \text{otherwise} \end{cases}$$

- $S = \sum_{i=1}^{N} v_i I_i$
-

$$\begin{aligned} E[S] &= \sum_{i=1}^{N} v_i E[I_i] \\ \mathrm{Var}(S) &= \sum_{i=1}^{N} \mathrm{Var}(v_i I_i) + 2 \sum \sum_{i<j} \mathrm{Cov}(v_i I_i, v_j I_j) \\ &= \sum_{i=1}^{N} v_i^2 \mathrm{Var}(I_i) + 2 \sum \sum_{i<j} v_i v_j \mathrm{Cov}(I_i, I_j) \end{aligned}$$

- $E[I_i] = \frac{n}{N}$
- $E[I_i I_j] = \frac{n}{N} \frac{n-1}{N-1}$
-

$$\begin{aligned} \mathrm{Var}(I_i) &= \frac{n}{N}\left(1 - \frac{n}{N}\right) \\ \mathrm{Cov}(I_i, I_j) &= \frac{n(n-1)}{N^2(N-1)} - \left(\frac{n}{N}\right)^2 \\ &= \frac{-n(N-n)}{N^2(N-1)} \end{aligned}$$

- Hence

$$\begin{aligned} E[S] &= n \sum_{i=1}^{N} \frac{v_i}{N} = n\overline{v} \\ \mathrm{Var}(S) &= \frac{n}{N}\left(\frac{N-n}{N}\right) \sum_{i=1}^{N} v_i^2 - \frac{2n(N-n)}{N^2(N-1)} \sum \sum_{i<j} v_i v_j \end{aligned}$$

- $\text{Var}(S) = \frac{n(N-n)}{N-1}\left(\frac{\Sigma_{i=1}^{N} v_i^2}{N} - \overline{v}^2\right)$
- $E[S] = n\overline{v} = np \quad \text{since} \quad \overline{v} = \frac{Np}{N} = p$
- $$\begin{aligned} \text{Var}(S) &= \frac{n(N-n)}{N-1}\left(\frac{Np}{N} - p^2\right) \\ &= \frac{n(N-n)}{N-1} p(1-p) \end{aligned}$$
- $E\left[\frac{S}{n}\right] = p$
- $\text{Var}\left(\frac{S}{n}\right) = \frac{N-n}{n(N-1)} p(1-p)$

Correlation:

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

- $0 \le \text{Var}\left(\frac{X}{\sigma_x} + \frac{Y}{\sigma_y}\right)$ implies $-1 \le \rho(X,Y)$.
- $0 \le \text{Var}\left(\frac{X}{\sigma_x} - \frac{Y}{\sigma_y}\right)$ implies $1 \ge \rho(X,Y)$.
- If $\rho(X,Y) = 1$, then $Y = a + bX$ where $b = \sigma_y/\sigma_x$.

- If $\rho(X,Y) = -1$, then $Y = a + bX$ where $b = -\sigma_y/\sigma_x$.
- X and Y are **uncorrelated** if $\rho(X,Y) = 0$.

Example 7.3e. 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

- $E[I_A] = P(A)$
- $E[I_B] = P(B)$
- $E[I_A I_B] = P(AB)$

$$\begin{aligned} \mathrm{Cov}(I_A, I_B) &= P(AB) - P(A)P(B) \\ &= P(B)[P(A|B) - P(A)] \end{aligned}$$

- 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 greater than, equal to, or less than $P(A)$.

Example 7.3f. Let $X_1, \dots, X_n$ be independent and identically distributed random variables having variance σ^2. Show that

$$\mathrm{Cov}(X_i - \overline{X}, \overline{X}) = 0$$

-
$$\begin{aligned}
\mathrm{Cov}(X_i - \overline{X}, \overline{X}) &= \mathrm{Cov}(X_i, \overline{X}) - \mathrm{Cov}(\overline{X}, \overline{X}) \\
&= \mathrm{Cov}\left(X_i, \frac{1}{n}\sum_{j=1}^{n} X_j\right) - \mathrm{Var}(\overline{X}) \\
&= \frac{1}{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) - \frac{\sigma^2}{n} \\
&= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0
\end{aligned}$$

-
$$\mathrm{Cov}(X_i, X_j) = \begin{cases} 0 & \text{if } j \neq i \text{ by independence} \\ \sigma^2 & \text{if } j = i \text{ since } \mathrm{Var}(X_i) = \sigma^2 \end{cases}$$

- Although $\overline{X}$ and the deviation $X_i - \overline{X}$ are uncorrelated, they are not, in general, independent.
- If X_i's are $N(\mu, \sigma^2)$, then $\overline{X}$ and the deviation $(X_i - \overline{X})$'s are independent.

Example 7.3g. Consider m independent trials, each of which results in any of r possible outcomes with probabilities $P_1, P_2, \ldots, P_r, \sum_1^r P_i = 1$.

- N_i: Denote the number of the m trials that result in outcome i, then $N_1, N_2, \ldots, N_r$ have the multinomial distribution

$$P\{N_1 = n_1, N_2 = n_2, \ldots, N_r = n_r\}$$

$$= \frac{m!}{n_1!n_2! \ldots, n_r!} P_1^{n_1} P_2^{n_2} \cdots P_r^{n_r} \qquad \sum_{i=1}^{r} n_i = m$$

- For $i \neq j$ it seems likely that when N_i is large N_j would tend to be small, and hence it is intuitive that they should be negatively correlated.

- Let us compute their covariance by using Proposition 3.2(iv) and the representation

$$N_i = \sum_{k=1}^{m} I_i(k) \quad \text{and} \quad N_j = \sum_{k=1}^{m} I_j(k)$$

where

$$I_i(k) = \begin{cases} 1 & \text{if trial } k \text{ results in outcome } i \\ 0 & \text{otherwise} \end{cases}$$

$$I_j(k) = \begin{cases} 1 & \text{if trial } k \text{ results in outcome } j \\ 0 & \text{otherwise} \end{cases}$$

- From Proposition 3.2(iv) we have

$$\mathrm{Cov}(N_i, N_j) = \sum_{\ell=1}^{m} \sum_{k=1}^{m} \mathrm{Cov}(I_i(k), I_j(\ell))$$

- Now, when $k \neq \ell$,

$$\mathrm{Cov}(I_i(k), I_j(\ell)) = 0$$

since the outcome of trial k is independent of the outcome of trial ℓ.

- On the other hand,

$$\begin{aligned} \mathrm{Cov}(I_i(\ell), I_j(\ell)) &= E[I_i(\ell)I_j(\ell)] - E[I_i(\ell)]E[I_j(\ell)] \\ &= 0 - P_iP_j = -P_iP_j \end{aligned}$$

where the above uses that $I_i(\ell)I_j(\ell) = 0$ since trial ℓ cannot result in both outcome i and outcome j.

- Hence we obtain that

$$\mathrm{Cov}(N_i, N_j) = -mP_iP_j$$

which is in accord with our intuition that N_i and N_j are negatively correlated.

7.4 Conditional expectation

7.4.1 Definitions

Discrete case:

$$p_{X|Y}(x|y) = P\{X = x \mid Y = y\} = \frac{p(x,y)}{p_Y(y)}$$

$$E[X = x \mid Y = y] = \sum_x xP\{X = x \mid Y = y\}$$

Example 7.4a. If X and Y are independent binomial random variables with identical parameters n and p, calculate the conditional expected value of X, given that $X + Y = m$.

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

- The conditional distribution of X, given that $X+Y=m$, is the hypergeometric distribution $(2n, n, m)$.

- $E[X \mid X+Y=m] = m/2$

Continuous case:

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\ E[X \mid Y=y] &= \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx \end{aligned}$$

Example 7.4b. Suppose that the joint density of X and Y is given by

$$f(x,y) = \frac{e^{-x/y}e^{-y}}{y} \quad 0 < x, y < \infty$$

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

- $$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\ &= \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dx} \\ &= \frac{(1/y)e^{-x/y}e^{-y}}{\int_0^{\infty}(1/y)e^{-x/y}e^{-y}dx} \\ &= \frac{(1/y)e^{-x/y}}{\int_0^{\infty}(1/y)e^{-x/y}dx} \\ &= \frac{1}{y}e^{-x/y} \end{aligned}$$

- The conditional distribution of X, given that $Y = y$, is $\exp(1/y)$.

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

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

$$E[\sum_{i=1}^{n} X_i | Y=y] = \sum_{i=1}^{n} E[X_i|Y=y]$$

7.4.2 Computing expectations by conditioning

Proposition 4.1:

$$E[X] = E[E[X|Y]] \tag{4.1}$$

$$E[X] = \begin{cases} \sum_y E[X|Y=y] P\{Y=y\} & \text{discrete case} \\ \int_{-\infty}^{\infty} E[X|Y=y] f_Y(y)\, dy & \text{continuous case} \end{cases}$$

This is an extremely useful result that often enables us to easily compute expectations by first conditioning on some appropriate random variable.

Example 7.4c. 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. If we assume that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until he reaches safety?

- X: The amount of time until the miner reaches safety.
- Y: The door he initially chooses.

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

- Note that

$$\begin{aligned} E[X|Y=1] &= 3 \\ E[X|Y=2] &= 5 + E[X] \\ E[X|Y=3] &= 7 + E[X] \end{aligned} \tag{4.3}$$

- $E[X] = \frac{1}{3}(3 + 5 + E[X] + 7 + E[X])$

- $E[X] = 15$

Example 7.4d. *Expectation of a random number of random variables.* Suppose that the number of people entering a department store on a given day is a random variable with mean 50. Suppose further that amounts of money spent by these customers are independent random variables having a common mean of 8. Assume also that the amount of money spent by a customer is also independent of the total number of customers to enter the store. What is the expected amount of money spent in the store in a given day?

- N: The number of customers that enter the store.
- X_i: The amount spent by the ith such customer.

$$E\left[\sum_1^N X_i\right] = E\left[E\left[\sum_1^N X_i | N\right]\right]$$

- $$\begin{aligned} E\left[\sum_1^N X_i | N = n\right] &= E\left[\sum_1^n X_i | N = n\right] \\ &= E\left[\sum_1^n X_i\right] \quad \text{by the independence of the } X_i \text{ and } N \\ &= nE[X] \quad \text{where } E[X] = E[X_i] \end{aligned}$$
- $E\left[\sum_1^N X_i | N\right] = NE[X]$
- Thus
$$E\left[\sum_{i=1}^N X_i\right] = E[NE[X]] = E[N]E[X]$$
- The expected amount of money spent: $50 \times 8 = 400$

Example 7.4e. Consider n points that are independently and uniformly distributed on the interval (0,1). Say that any one of these points is "isolated" if there are no other points within a distance d of it, where d is a specified constant such that $0 < d < \frac{1}{2}$. Compute the expected number of the n points that are isolated from the others.

- Let the points be $U_1, \ldots, U_n$, and define I_j

as the indicator variable for the event that U_j is an isolated point.

- $E\left[\sum_{j=1}^{n} I_j\right] = \sum_{j=1}^{n} E[I_j]$
- $E[I_j] = \int_0^1 E[I_j|U_j = x]\, dx$
-

$$\begin{aligned}
E[I_j] &= \int_0^d E[I_j|U_j = x]\, dx + \int_d^{1-d} E[I_j|U_j = x]\, dx + \int_{1-d}^1 E[I_j|U_j = x]\, dx \\
&= \int_0^d (1-d-x)^{n-1}\, dx + \int_d^{1-d} (1-2d)^{n-1}\, dx \\
&\quad + \int_{1-d}^1 (1-x+d)^{n-1}\, dx \\
&= \int_{1-2d}^{1-d} y^{n-1}\, dy + (1-2d)(1-2d)^{n-1} + \int_d^{2d} y^{n-1}\, dy \\
&= \frac{(1-d)^n}{n} - \frac{(1-2d)^n}{n} + (1-2d)^n + \frac{(2d)^n}{n} - \frac{d^n}{n}
\end{aligned}$$

- $E\left[\sum_{j=1}^{n} I_j\right] = (1-d)^n + (n-1)(1-2d)^n + (2^n - 1)d^n$
- If $d = c/n$, $E\left[\sum_{j=1}^{n} I_j\right] \approx e^{-c} + (n-1)e^{-2c}$

Example 7.4f. An urn contains a white and b black balls. One ball at a time is randomly withdrawn until the first white ball is drawn.

Find the expected number of black balls that are withdrawn.

- X: The number of black balls withdrawn.

$$Y = \begin{cases} 1 & \text{if the first ball selected is white} \\ 0 & \text{if the first ball selected is black} \end{cases}$$

- $M_{a,b} = E[X] = E[X|Y = 1]P\{Y = 1\} + E[X|Y = 0]P\{Y = 0\}$
- $E[X|Y = 1] = 0$
- $E[X|Y = 0] = 1 + M_{a,b-1}$
- Since $P\{Y = 0\} = b/(a + b)$, we see that

$$M_{a,b} = \frac{b}{a+b}[1 + M_{a,b-1}]$$

- $M_{a,0}$ is clearly equal to 0,

$$\begin{aligned} M_{a,1} &= \frac{1}{a+1}[1 + M_{a,0}] = \frac{1}{a+1} \\ M_{a,2} &= \frac{2}{a+2}[1 + M_{a,1}] = \frac{2}{a+2}\left[1 + \frac{1}{a+1}\right] = \frac{2}{a+1} \\ M_{a,3} &= \frac{3}{a+3}[1 + M_{a,2}] = \frac{3}{a+3}\left[1 + \frac{2}{a+1}\right] = \frac{3}{a+1} \end{aligned}$$

-

- $M_{a,b} = \frac{b}{a+1}$

Example 7.4g. *Variance of the geometric distribution.* Independent trials each resulting in a success with probability p are successively performed. Let N be the time of the first success. Find $\mathrm{Var}(N)$.

- $\mathrm{Var}(N) = E[N^2] - (E[N])^2$
- However,

$$
\begin{aligned}
E[N^2|Y = 1] &= 1 \\
E[N^2|Y = 0] &= E[(1 + N)^2]
\end{aligned}
$$

-

$$
\begin{aligned}
E[N^2] &= E[N^2|Y = 1]P\{Y = 1\} + E[N^2|Y = 0]P\{Y = 0\} \\
&= p + (1 - p)E[(1 + N)^2] \\
&= 1 + (1 - p)E[2N + N^2]
\end{aligned}
$$

- $E[N^2] = 1 + \frac{2(1-p)}{p} + (1 - p)E[N^2]$
- $E[N^2] = \frac{2-p}{p^2}$

- Therefore,

$$\begin{aligned}\mathrm{Var}(N) &= E[N^2] - (E[N])^2 \\ &= \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 \\ &= \frac{1-p}{p^2}\end{aligned}$$

Example 7.4h. Let $U_1, U_2, \ldots$ be a sequence of independent uniform (0,1) random variables. Find $E[N]$ when

$$N = \min\left\{n : \sum_{i=1}^{n} U_i > 1\right\}$$

- $N(x) = \min\left\{n : \sum_{i=1}^{n} U_i > x\right\}$
- $m(x) = E[N(x)]$
- $m(x) = \int_0^1 E[N(x)|U_1 = y]\,dy$
-

$$E[N(x)|U_1 = y] = \begin{cases} 1 & \text{if } y > x \\ 1 + m(x-y) & \text{if } y \le x \end{cases}$$

-
$$\begin{aligned} m(x) &= 1 + \int_0^x m(x-y)\,dy \\ &= 1 + \int_0^x m(u)du \quad \text{by letting} \quad u = x - y \end{aligned}$$
- $m'(x) = m(x)$
- $\frac{m'(x)}{m(x)} = 1$
- $\log[m(x)] = x + c$
- $m(x) = ke^x$
- Since $m(0) = 1$ we see that $k = 1$, then
$$m(x) = e^x$$

7.4.3 Computing probabilities by conditioning

$$\begin{aligned} X &= \begin{cases} 1 & \text{if } E \text{ occurs} \\ 0 & \text{if } E \text{ does not occur} \end{cases} \\ E[X] &= P(E) \\ E[X|Y=y] &= P(E|Y=y) \\ P(E) &= \sum_y P(E|Y=y)P(Y=y) \quad \text{if } Y \text{ is discrete} \\ &= \int_{-\infty}^{\infty} P(E|Y=y)f_Y(y)\,dy \quad \text{if } Y \text{ is continuous} \end{aligned}$$

Example 7.4i. *The best prize problem.* Suppose that we are to be presented with n distinct prizes in sequence. After being presented with a prize we must immediately decide whether to accept it or to reject it and consider the next prize. The only information we are given when deciding whether to accept a prize is the relative rank of that prize compared to ones already seen. That is, for instance, when the fifth prize is presented, we learn how it compares with the four prizes already seen. Suppose that once a prize is rejected it is list, and that our objective is to maximize the probability of obtaining the best prize. Assuming that all $n!$ orderings of the prizes are equally likely, how well can we do?

- Consider the strategy that rejects the first k prizes and then accepts the first one that is better than all of those first k.

$$P_k(\text{best}) = \sum_{i=1}^{n} P_k(\text{best}|X = i)P(X = i)$$

$$= \frac{1}{n} \sum_{i=1}^{n} P_k(\text{best}|X = i)$$

- $P_k(\text{best}|X = i) = 0 \quad i \leq k$
- $$\begin{aligned} P_k(\text{best}) &= \frac{k}{n} \sum_{i=k+1}^{n} \frac{1}{i-1} \\ &\approx \frac{k}{n} \int_{k+1}^{n} \frac{1}{x+1}\, dx \\ &= \frac{k}{n} \log\left(\frac{n-1}{k}\right) \\ &\approx \frac{k}{n} \log\left(\frac{n}{k}\right) \end{aligned}$$
- Now, if we consider the function
$$g(x) = \frac{x}{n} \log\left(\frac{n}{x}\right)$$
- $$g'(x) = \frac{1}{n} \log\left(\frac{n}{x}\right) - \frac{1}{n}$$
- $$g'(x) = 0 \Rightarrow \log\left(\frac{n}{x}\right) = 1 \Rightarrow x = \frac{n}{e}$$

Example 7.4j. Let U be a uniform random variable on (0,1), and suppose that the conditional distribution of X, given that $U = p$, is binomial with parameters n and p. Find the probability mass function of X.

- Conditioning on the value of U:

$$\begin{aligned} P\{X = i\} &= \int_0^1 P\{X = i|U = p\} f_U(p)\, dp \\ &= \int_0^1 P\{X = i|U = p\}\, dp \\ &= \frac{n!}{i!(n-i)!} \int_0^1 p^i (1-p)^{n-i}\, dp \end{aligned}$$

- $\int_0^1 p^i (1-p)^{n-i} dp = \dfrac{i!(n-i)!}{(n+1)!}$
- Hence we obtain that

$$P\{X = i\} = \frac{1}{n+1} \quad i = 0, \ldots, n$$

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

- Another argument:
  - $U, U_1, \ldots, U_n$ are independent uniform$(0, 1)$.
  - X: The number of the random variables $U_1, \ldots, U_n$ that are smaller than U.
  - Since all the random variables $U, U_1, \ldots, U_n$ have the same distribution, it follows that U is equally likely to be the smallest, or the second smallest, or the largest of them; so X is equally likely to be any of the values $0, 1, \ldots, n$.

Example 7.4k. Suppose that X and Y are independent continuous random variables having densities f_X and f_Y, respectively. Compute $P\{X < Y\}$.

- Conditioning on the value of Y:

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

where

$$F_X(y) = \int_{-\infty}^{y} f_X(X)\, dx$$

- Special case: If $f_X = f_Y$, then $P(X < Y) = \frac{1}{2}$.

Example 7.4l. Suppose that X and Y are independent continuous random variables. Find the distribution of $X + Y$.

- Conditioning on the value of Y:

$$\begin{aligned} P\{X+Y<a\} &= \int_{-\infty}^{\infty} P\{X+Y<a|Y=y\} f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} P\{X+y<a|Y=y\} f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} P\{X<a-y\} f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} F_X(a-y) f_Y(y)\, dy \end{aligned}$$

7.4.4 Conditional variance

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

Proposition 4.2: The conditional variance formula

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])$$

- $\text{Var}(X|Y) = E[X^2|Y] - (E[X|Y])^2$
- $E[\text{Var}(X|Y)] = E[X^2] - E[(E[X|Y])^2]$
- $\text{Var}(E[X|Y]) = E[(E[X|Y])^2] - (E[X])^2$

Example 7.4m. 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 is the mean and variance of the number of passengers that enter the train?

- $N(t)$: The number of arrivals by t.
- Y: The time at which the train arrives.
- The random variable of interest is $N(t)$.

$$E[N(Y)|Y = t] = E[N(t)|Y = t]$$

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

- $E[N(Y)|Y] = \lambda Y$
- $E[N(Y)] = \lambda E[Y] = \frac{\lambda T}{2}$
-
$$
\begin{aligned}
\mathrm{Var}(N(Y)|Y=t) &= \mathrm{Var}(N(t)|Y=t) \\
&= \mathrm{Var}(N(t)) \quad \text{by independence} \\
&= \lambda t \\
\mathrm{Var}(N(Y)|Y) &= \lambda Y \\
E[N(Y)|Y] &= \lambda Y
\end{aligned}
$$
- From the conditional variance formula:
$$
\begin{aligned}
\mathrm{Var}(N(Y)) &= E[\lambda Y] + \mathrm{Var}(\lambda Y) \\
&= \lambda \frac{T}{2} + \lambda^2 \frac{T^2}{12}
\end{aligned}
$$

Example 7.4n. *Variance of a random number of random variables.* Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables and let N be a non-negative integer-valued random variable that is independent of the sequence X_i, $i \geq 1$. To compute $\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right)$, we condition on N:

$$
\begin{aligned}
E\left[\sum_{i=1}^{N} X_i|N\right] &= NE[X] \\
\mathrm{Var}\left(\sum_{i=1}^{N} X_i|N\right) &= N\mathrm{Var}(X) \\
\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) &= E[N]\mathrm{Var}(X) + (E[X])^2\mathrm{Var}(N)
\end{aligned}
$$

7.5 Conditional expectation and prediction

- $X = x$ is observed.
- Use $g(x)$ to predict Y.
- Choose g so as to min $E[(Y - g(X))^2]$.

Proposition 7.5.1:

$$E[(Y - g(X))^2] \geq E[(Y - E[Y|X])^2]$$

- $E[(Y - g(X))^2|X] = E[(Y - E[Y|X] + E[Y|X] - g(X))^2|X] =$

$E[(Y-E[Y|X])^2|X]+E[(E[Y|X]-g(X))^2|X]+$
$2E[(Y-E[Y|X])(E[Y|X]-g(X))|X]$

- $E[(Y-E[Y|X])(E[Y|X]-g(X))|X]=0$
- $E[(Y-g(X))^2|X] \geq E[(Y-E[Y|X])^2|X]$

Example 7.5a. Suppose that the son of a man of height x (in inches) attains a height that is normally distributed with mean $x+1$ and variance 4. What is the best prediction of the height at full growth of the son of a man who is 6 feet tall?

- $Y = X + 1 + e$ where $e \sim N(0, 4)$.

$$\begin{aligned} E[Y|X=72] &= E[X+1+e|X=72] \\ &= 73 + E[e|X=72] \\ &= 73 + E(e) \quad \text{by independence} \\ &= 73 \end{aligned}$$

Example 7.5b. Suppose that if a signal value s is sent from location A, then the signal value received at location B is normally distributed with parameters $(s, 1)$. If S, the value of the signal sent at A, is normally distributed with parameters (μ, σ^2), what is the best estimate of the signal sent if R, the value received at B, is equal to r?

•

$$\begin{aligned} f_{S|R}(s|r) &= \frac{f_{S,R}(s,r)}{f_R(r)} \\ &= \frac{f_S(s) f_{R|S}(r|s)}{f_R(r)} \\ &= K e^{-(s-\mu)^2/2\sigma^2} e^{-(r-s)^2/2} \end{aligned}$$

$$\begin{aligned} \frac{(s-\mu)^2}{2\sigma^2} + \frac{(r-s)^2}{2} &= s^2 \left(\frac{1}{2\sigma^2} + \frac{1}{2} \right) - \left(\frac{\mu}{\sigma^2} + r \right) s + C_1 \\ &= \frac{1+\sigma^2}{2\sigma^2} \left[s^2 - 2 \left(\frac{\mu + r\sigma^2}{1+\sigma^2} \right) s \right] + C_1 \\ &= \frac{1+\sigma^2}{2\sigma^2} \left(s - \frac{\mu + r\sigma^2}{1+\sigma^2} \right)^2 + C_2 \end{aligned}$$

where C_1 and C_2 do not depend on s.

- Hence

$$f_{S|R}(s|r) = C \exp\left\{ \frac{-\left[s - \frac{(\mu + r\sigma^2)}{1+\sigma^2}\right]^2}{2\left(\frac{\sigma^2}{1+\sigma^2}\right)} \right\}$$

- $E[S|R = r] = \frac{\mu + r\sigma^2}{1+\sigma^2}$
- $\mathrm{Var}(S|R = r) = \frac{\sigma^2}{1+\sigma^2}$
- $E[S|R = r] = \frac{1}{1+\sigma^2}\mu + \frac{\sigma^2}{1+\sigma^2}r$

Example 7.5c. In digital signal processing raw continuous analog data X must be quantized, or discretized, in order to obtain a digital representation. In order to quantize the raw data X, an increasing set of numbers $a_i, i = 0, \pm 1, \pm 2, \ldots$, such that $\lim_{i \to +\infty} a_i = \infty$, $\lim_{i \to -\infty} a_i = -\infty$, is fixed and the raw data are then quantized according to the interval $(a_i, a_{i+1}]$ in which X lies. Let us denote by y_i the discretized value when $X \in (a_i, a_{i+1}]$, and let Y denote the observed discretized value-that is,

$$Y = y_i \quad \text{if } a_i < X \leq a_{i+1}$$

The distribution of Y is given by

$$P\{Y = y_i\} = F_X(a_{i+1}) - F_X(a_i)$$

Suppose now that we want to choose the values $y_i, i = 0, \pm 1, \pm 2, \ldots$ so as to minimize $E[(X-Y)^2]$, the expected mean square difference between the raw data and their quantized version.

(a) Find the optimal values $y_i, i = 0, \pm 1, \ldots$
For the optimal quantizer Y show that:

(b) $E[X] = E[Y]$, so the mean square error quantizer preserves the input mean;

(c) $\text{Var}(Y) = \text{Var}(X) - E[(X-Y)^2]$.

- (a)
- $E[(X-Y)^2] = \sum_i E[(X-y_i)^2 | a_i < X \leq a_{i+1}] P\{a_i < X \leq a_{i+1}\}$

- $I = i \quad \text{if } a_i < X \leq a_{i+1}$
- Then
  $E[(X - y_i)^2 | a_i < X \leq a_{i+1}] = E[(X - y_i)^2 | I = i]$
-
$$\begin{aligned} y_i &= E[X|I=i] \\ &= E[X|a_i < X \leq a_{i+1}] \\ &= \int_{a_i}^{a_{i+1}} \frac{x f_X(s)\, dx}{F_X(a_{i+1}) - F_X(a_i)} \end{aligned}$$
- (b) $E[Y] = E[X]$
- (c)
$$\begin{aligned} \mathrm{Var}(X) &= E[\mathrm{Var}(X|I)] + \mathrm{Var}(E[X|I]) \\ &= E[E[(X-Y)^2|I]] + \mathrm{Var}(Y) \\ &= E[(X-Y)^2] + \mathrm{Var}(Y) \end{aligned}$$

Best linear predictor of Y w.r.t X

$$\min_{a,b} E[(Y - (a + bX))^2]$$

- $\frac{\partial}{\partial a} E[(Y - (a + bX))^2] = -2E[Y] + 2a + 2bE[X]$

- $\frac{\partial}{\partial b} E[(Y-(a+bX))^2] = -2E[XY]+2aE[X]+2bE[X^2]$
- $b = \frac{E[XY]-E[X]E[Y]}{E[X^2]-(E[X])^2} = \frac{\text{Cov}(X,Y)}{\sigma_x^2} = \rho\frac{\sigma_y}{\sigma_x}$
- $a = E[Y] - bE[X]$
- Best linear predictor of Y w.r.t. X

$$\mu_y + \frac{\rho\sigma_y}{\sigma_x}(X - \mu_x)$$

- Mean square error of this predictor:

$$E\left[\left(Y - \mu_y - \frac{\rho\sigma_y}{\sigma_x}(X - \mu_x)\right)^2\right] = \sigma_y^2(1-\rho^2)$$

Example 7.5d. An example in which the conditional expectation of Y given X is linear in X, and hence the best linear predictor of Y with respect to X is the best overall predictor, is when X and Y have a bivariate normal distribution. In this case their joint density is given by

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}$$

- $$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left(y-\mu_y-\frac{\rho\sigma_y}{\sigma_x}(x-\mu_x)\right)^2\right\}$$
- $E[Y|X=x] = \mu_y + \frac{\rho\sigma_y}{\sigma_x}(x-\mu_x)$
- $\mathrm{Var}(Y|X=x) = \sigma_y^2(1-\rho^2)$

7.6 Moment generating functions

$$\begin{aligned} M(t) &= E[e^{tX}] \\ &= \begin{cases} \sum_x e^{tx}p(x) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\,dx & \text{if } X \text{ is continuous} \end{cases} \end{aligned}$$

$M'(t) = E[Xe^{tX}]$	$M'(0) = E[X]$
$M''(t) = E[X^2e^{tX}]$	$M''(0) = E[X^2]$
$M^{(n)}(t) = E[X^ne^{tX}]$	$M^{(n)}(0) = E[X^n]$

Example 7.6a. *Binomial distribution with parameters n and p.* If X is a binomial random variable with parameters n and p, then

$$M(t) = E[e^{tX}]$$

$$
\begin{aligned}
&= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\
&= (pe^t + 1 - p)^n
\end{aligned}
$$

- $M'(t) = n(pe^t + 1 - p)^{n-2} pe^t$ and $E[X] = M'(0) = np$.
- $M''(t) = n(n-1)(pe^t + 1 - p)^{n-2}(pe^t)^2 + n(pe^t + 1 - p)^{n-1} pe^t$ and $E[X^2] = M''(0) = n(n-1)p^2 + np$.
- The variance of X is given by

$$
\begin{aligned}
\text{Var}(X) &= E[X^2] - (E[X])^2 \\
&= n(n-1)p^2 + np - n^2p^2 \\
&= np(1-p)
\end{aligned}
$$

Example 7.6b. *Poisson distribution with mean λ.* If X is a Poisson random variable with parameter λ, then

$$M(t) = E[e^{tX}]$$

$$
\begin{aligned}
&= \sum_{n=0}^{\infty} \frac{e^{tn} e^{-\lambda} \lambda^n}{n!} \\
&= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} \\
&= e^{-\lambda} e^{\lambda e^t} \\
&= \exp\{\lambda(e^t - 1)\}
\end{aligned}
$$

- Differentiation yields

$$
\begin{aligned}
M'(t) &= \lambda e^t \exp\{\lambda(e^t - 1)\} \\
M''(t) &= (\lambda e^t)^2 \exp\{\lambda(e^t - 1)\} + \lambda e^t \exp\{\lambda(e^t - 1)\}
\end{aligned}
$$

- Thus

$$
\begin{aligned}
E[X] &= M'(0) = \lambda \\
E[X^2] &= M''(0) = \lambda^2 + \lambda \\
\mathrm{Var}(X) &= E[X^2] - (E[X])^2 \\
&= \lambda
\end{aligned}
$$

- Hence both the mean and the variance of the Poisson random variable equal λ.

Example 7.6c. *Exponential distribution with parameter λ*

$$
M(t) = E[e^{tX}]
$$

$$
\begin{aligned}
&= \int_0^\infty e^{tX} \lambda e^{-\lambda x}\, dx \\
&= \lambda \int_0^\infty e^{-(\lambda - t)x}\, dx \\
&= \frac{\lambda}{\lambda - t} \quad \text{for } t < \lambda
\end{aligned}
$$

- Differentiation of $M(t)$ yields

$$M'(t) = \frac{\lambda}{(\lambda - t)^2} \quad M''(t) = \frac{2\lambda}{(\lambda - t)^3}$$

- Hence

$$E[X] = M'(0) = \frac{1}{\lambda} \quad E[X^2] = M''(0) = \frac{2}{\lambda^2}$$

- The variance of X is given by

$$
\begin{aligned}
\mathrm{Var}(X) &= E[X^2] - (E[X])^2 \\
&= \frac{1}{\lambda^2}
\end{aligned}
$$

Example 7.6d. *Normal distribution.* We first compute the moment generating function of a unit normal random variable with parameters 0 and 1.

- Letting Z be such a random variable,

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

- $M_Z(t) = e^{t^2/2}$
- $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$

$$\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= E[e^{t(\mu+\sigma Z)}] \\
&= E[e^{t\mu} e^{t\sigma Z}] \\
&= e^{t\mu} M_Z(t\sigma) \\
&= e^{t\mu} e^{(t\sigma)^2/2}
\end{aligned}$$

$$= \exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\}$$

- By differentiating, we obtain

$$M_X'(t) = (\mu + t\sigma^2)\exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\}$$
$$M_X''(t) = (\mu + t\sigma^2)^2\exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\} + \sigma^2\exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\}$$

- Thus

$$E[X] = M'(0) = \mu$$
$$E[X^2] = M''(0) = \mu^2 + \sigma^2$$

implying that

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

Suppose that X and Y are independent and have moment generating functions $M_X(t)$ and $M_Y(t)$, respectively. Then

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

Uniqueness of moment generating function: If $M_X(t)$ exists and is finite in some region about $t = 0$, then the distribution of X is uniquely determined.

For example, if $M_X(t) = (1/2)^{10}(e^t + 1)^{10}$, then X is a binomial$(10, 1/2)$.

Example 7.6e. Suppose that the moment generating function of a random variable X is given by $M(t) = e^{3(e^t - 1)}$. What is $P\{X = 0\}$?

- $M(t)$ is the moment generating function of a Poisson random variable with mean 3.
- $P\{X = 0\} = e^{-3}$

Example 7.6f. *Sums of independent binomial random variables.* If X and Y are independent binomial random variables with parameters (n, p) and (m, p), respectively, what is the distribution of $X + Y$?

- The moment generating function of $X + Y$

is given by

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)M_Y(t) \\ &= (pe^t + 1 - p)^n (pe^t + 1 - p)^m \\ &= (pe^t + 1 - p)^{m+n} \end{aligned}$$

- Thus $X + Y$ is binomial distributed with parameters $m + n$ and p.

Example 7.6g. *Sums of independent Poisson random variables.* Calculate the distribution of $X + Y$ when X and Y are independent Poisson random variables with means λ_1 and λ_2, respectively.

-

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)M_Y(t) \\ &= \exp\{\lambda_1(e^t - 1)\} \exp\{\lambda_2(e^t - 1)\} \\ &= \exp\{(\lambda_1 + \lambda_2)(e^t - 1)\} \end{aligned}$$

- Thus $X + Y$ is Poisson distributed with parameters $\lambda_1 + \lambda_2$.

Example 7.6h. *Sums of independent normal random variables.* Show that if X and Y are independent normal random variables with parameters (μ_1, σ_1^2) and (μ_2, σ_2^2), respectively, then $X+Y$ is normal with mean $\mu_1+\mu_2$ and variance $\sigma_1^2+\sigma_2^2$.

- $$\begin{aligned} M_{X+Y}(t) &= M_X(t)M_Y(t) \\ &= \exp\left\{\frac{\sigma_1^2 t^2}{2} + \mu_1 t\right\} \exp\left\{\frac{\sigma_2^2 t^2}{2} + \mu_2 t\right\} \\ &= \exp\left\{\frac{(\sigma_1^2+\sigma_2^2)t^2}{2} + (\mu_1+\mu_2)t\right\} \end{aligned}$$

Example 7.6i. Compute the moment generating function of a chi-squared random variable with n degrees of freedom.

- We can represent such a random variable as

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

- $M(t) = (E[e^{tZ^2}])^n$ where Z is a standard normal.

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

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

Example 7.6j. *Moment generating function of the sum of a random number of random variables.* Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, and let N be a nonnegative, integer-valued random variable that is independent of the sequence X_i, $i \geq 1$. We want to compute the moment generating function of

$$Y = \sum_{i=1}^{N} X_i$$

- Condition on N

$$E[\exp\{t \sum_1^N X_i\} | N = n] = E[\exp\{t \sum_1^n X_i\} | N = n]$$

$$
\begin{aligned}
&= E[\exp\{t\sum_1^n X_i\}] \\
&= [M_X(t)]^n
\end{aligned}
$$

where

$$M_X(t) = E[e^{tX_i}]$$

- $E[e^{tY}|N] = (M_X(t))^N$
- $M'_Y(t) = E[N(M_X(t))^{N-1}M'_X(t)]$

$$
\begin{aligned}
E[Y] &= M'_Y(0) \\
&= E[N(M_X(0))^{N-1}M'_X(0)] \\
&= E[NE[X]] \\
&= E[N]E[X]
\end{aligned}
$$

-

$$
\begin{aligned}
E[Y^2] &= M''_Y(0) \\
&= E[N(N-1)(E[X])^2 + NE[X^2]] \\
&= (E[X])^2(E[N^2] - E[N]) + E[N]E[X^2] \\
&= E[N](E[X^2] - (E[X])^2) + (E[X])^2E[N^2] \\
&= E[N]\text{Var}(X) + (E[X])^2E[N^2]
\end{aligned}
$$

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

Example 7.6k. Let Y denote a uniform random variable on $(0, 1)$, and suppose that conditional on $Y = p$, the random variable X has a binomial distribution with parameters n and p. In Example 7.4j we showed that X is equally likely to take on any of the values $0, 1, \ldots, n$. Establish this result by using moment generating functions.

- $E[e^{tX}|Y = p] = (pe^t + 1 - p)^n$

$$\begin{aligned} E[e^{tX}] &= \int_0^1 (pe^t + 1 - p)^n \, dp \\ &= \frac{1}{e^t - 1} \int_1^{e^t} y^n \, dy \\ &= \frac{1}{n+1} \frac{e^{t(n+1)} - 1}{e^t - 1} \\ &= \frac{1}{n+1} (1 + e^t + e^{2t} + \cdots + e^{nt}) \end{aligned}$$

- X is uniformly distributed on $0, 1, \ldots, n$.

7.6.1 Joint moment generating functions

- $M(t_1, \ldots, t_n) = E[e^{t_1X_1+\cdots+t_nX_n}]$
- $M_{X_i}(t) = E[e^{tX_i}] = M(0, \ldots, 0, t, 0, \ldots, 0)$
- If $X_1, \ldots, X_n$ are independent if and only if

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

Example 7.6l. Let X and Y be independent normal random variables, each with mean μ and variance σ^2. In Example 7.7a of Chap. 6 we showed that $X+Y$ and $X-Y$ are independent.

- Let us now establish this result by computing their joint moment generating function.

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

- But we recognize the preceding as the joint moment generating function of the sum of a normal random variables 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.

Example 7.6m. 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 respective means λp and $\lambda(1-p)$.

- X: The total number of events.
- X_c: The number of them that are counted.
- Condition on X:
  $E[e^{sX_c+t(X-X_c)}|X=n] = e^{tn}E[e^{(s-t)X_c}|X=n]$

$$= e^{tn}(pe^{s-t} + 1 - p)^n$$
$$= (pe^s + (1-p)e^t)^n$$

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

7.7 Additional properties of normal random variables

7.7.1 The multivariate normal distribution

- $Z_1, \ldots, Z_n$ are a set of n independent unit normal.
- For some constants a_{ij} and μ_i,

$$\begin{aligned} X_1 &= a_{11}Z_1 + \cdots + a_{1n}Z_n + \mu_1 \\ &\vdots \\ X_i &= a_{i1}Z_1 + \cdots + a_{in}Z_n + \mu_i \end{aligned}$$

$$\vdots$$

$$X_m = a_{m1}Z_1 + \cdots + a_{mn}Z_n + \mu_m$$

then the random variables $X_1, \ldots, X_m$ are said to have a multivariate normal distribution.

- X_i is a normal random variable with $E[X_i] = \mu_i$ and $\mathrm{Var}(X_i) = \sum_{j=1}^{n} a_{ij}^2$.
- $\sum_{i=1}^{m} t_i X_i$ is a normal random variable with $E[\sum_{i=1}^{m} t_i X_i] = \sum_{i=1}^{m} t_i \mu_i$ and $\mathrm{Var}\left(\sum_{i=1}^{m} t_i X_i\right) = \sum_{i=1}^{m} \sum_{j=1}^{m} t_i t_j \mathrm{Cov}(X_i, X_j)$.
- $M(t_1, \ldots, t_m) = \exp\left\{\sum_{i=1}^{m} t_i \mu_i + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} t_i t_j \mathrm{Cov}(X_i, X_j)\right\} = \exp\left\{t'\mu + \frac{t'\Sigma t}{2}\right\}$

7.7.2 The joint distribution of the sample mean and sample variance

Let $X_i \sim N(\mu, \sigma^2)$.

- $\bar{X} = \sum_{i=1}^{n} X_i/n \sim N(\mu, \sigma^2/n)$
- $\text{Cov}(\bar{X}, X_i - \bar{X}) = 0$ for $i = 1, \ldots, n$.
- $\bar{X}, X_1 - \bar{X}, \ldots, X_n - \bar{X}$ are all linear combinations of the independent standard normals $(X_i - \bar{X})/\sigma$.
- Consider $Y \sim N(\mu, \sigma^2/n)$ independent of X_i's.
- $Y, X_1 - \bar{X}, \ldots, X_n - \bar{X}$ also has a multivariate normal and has the same expected values and covariances as the random variables $\bar{X}, X_1 - \bar{X}, \ldots, X_n - \bar{X}$.
- Then $\bar{X}, X_1 - \bar{X}, \ldots, X_n - \bar{X}$ also has a multivariate normal.
- But since a multivariate normal distribution is determined completely by its expected values and covariances, we can conclude that $\bar{X}$ is independent of $X_i - \bar{X}$'s.
- $(n-1)S^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2$

- $\dfrac{(n-1)S^2}{\sigma^2} + \left(\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 = \sum_{i=1}^{n} \left(\dfrac{X_i-\mu}{\sigma}\right)^2$
- Use moment generating function.
- $(1-2t)^{-(n-1)/2}(1-2t)^{-1/2} = (1-2t)^{-n/2}$

Proposition 7.1: If $X_1, \ldots, X_n$ are independent and identically distributed normal random variables with mean μ and variance σ^2, then the sample mean $\overline{X}$ and sample variance S^2 are independent. $\overline{X}$ is a normal random variable with mean μ and variance σ^2/n; $(n-1)S^2/\sigma^2$ is a chi-squared random variable with $n-1$ degrees of freedom.

*7.8 General definition of expectation

- There exist random variables that are neither discrete nor continuous.
- $X \sim$ Bernoulli$(1/2)$ and $Y \sim$ uniform$[0,1]$.
- Then $W = X$ if $X = 1$ and Y if $X \neq 1$ is neither a discrete nor a continuous random

variable.

- In order to define the expectation of an arbitrary random variable, we require the notion of a Stieltjes integral.
- $a = x_0 < x_1 < x_2 < \cdots < x_n = b$

$$\begin{aligned}
\int_a^b g(x)\,dx &= \lim_{n\to\infty} \sum_{i=1}^{n} g(x_i)(x_i - x_{i-1}) \\
\int_a^b g(x)\,dF(x) &= \lim_{n\to\infty} \sum_{i=1}^{n} g(x_i)[F(x_i) - F(x_{i-1})] \\
\int_{-\infty}^{\infty} g(x)\,dF(x) &= \lim_{a\to-\infty, b\to\infty} \int_a^b g(x)dF(x) \\
\int_{-\infty}^{\infty} g(x)\,dF(x) &= \int_{-\infty}^{\infty} g^+(x)\,dF(x) - \int_{-\infty}^{\infty} g^-(x)\,dF(x) \\
E[X] &= \int_{-\infty}^{\infty} x dF(x)
\end{aligned}$$

- Use of Stieltjes integrals avoids the necessity of having to give separate statements of theorems for the continuous and the discrete cases.
- Stieltjes integrals are mainly of theoretical interest because they yield a compact way of defining and dealing with the properties of expectation.

Summary

- **Expectation**:
  - Discrete:
    $$E[g(X,Y)] = \sum_y \sum_x g(x,y)p(x,y)$$
  - Continuous:
    $$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$$
  - $E[X+Y] = E[X] + E[Y]$
  - $E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$
- **Covariance**:
  - $\text{Cov}(X,Y) = E[(X-E[X])(Y-E[Y])] = E[XY] - E[X]E[Y]$
  - $\text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \text{Cov}(X_i, X_j)$
  - $\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j)$
- **Correlation**:
  $$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$
- **Conditional expected value**:

– Discrete case:

$$E[X \mid Y = y] = \sum xP\{X = x \mid Y = y\}$$

– Continuous case:

$$E[X \mid Y = y] = \int xf_{X|Y}(x|y)dx$$

- $E[X] = E[E[X|Y]]$

– Discrete case:

$$E[X] = \sum_y P\{Y = y\}$$

– Continuous case:

$$E[X] = \int E[X \mid Y = y]f(y)dy$$

- **Conditional variance**:

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

- **Conditional variance formula**:

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

- **Moment generating function**: $M(t) = E[e^{tX}]$

– $E[X^n] = \frac{d^n}{dt^n}M(t)|_{t=0}$

- The moment generating function uniquely determines the distribution function of the random variable.
- The moment generation function of the sum of independent random variables is equal to the product of their moment generation function.

- If $X_1, \ldots, X_n$ are all linear combinations of a finite set of independent standard normal random variables, then they are said to have a **multivariate normal distibution**.
- If $X_1, \ldots, X_n$ are independent and identically distributed normal random variables, then their **sample mean** $\overline{X} = \Sigma X_i/n$ and **sample variance** $S^2 = \Sigma(X_i - \overline{X})^2/(n - 1)$ are independent.
  - $\overline{X}$ is a normal variable with mean μ and variance σ^2/n
  - $(n-1)S^2/\sigma^2$ is a chi-square random variable with $n - 1$ degrees of freedom.