Section 7.1 Properties of Expectations: Introduction

Recall that we have

$$\boxed{E[X] = \sum_x xp(x)}$$

in the discrete case and

$$\boxed{E[X] = \int_{-\infty}^{\infty} xf(x)dx}$$

in the continuous case.

COROLLARY: If $P\{a \leq X \leq b\} = 1$, then

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

Proof:

$$E[X] = \sum_{x:p(x)>0} xp(x) \geq a \sum_{x:p(x)>0} p(x) = a$$

Likewise for upper bound and continuous case. ■

Section 7.2 Expectation of Sums of Random Variables

PROPOSITION: If X and Y have joint p.m.f. $p(x, y)$, then

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

If X and Y have joint p.d.f. $f(x, y)$, then

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

EXAMPLE: An accident occurs at a point X that is randomly (i.e. uniformly) selected on a road of length L. An ambulance is located at a point Y on the same road independent of X (also uniform). Find the expected distance between the accident and ambulance.

EXAMPLE: An accident occurs at a point X that is randomly (i.e. uniformly) selected on a road of length L. An ambulance is located at a point Y on the same road independent of X (also uniform). Find the expected distance between the accident and ambulance.

Solution: We need to find $E[|X-Y|]$. We are given that both X and Y are uniform$(0, L)$ and that they are independent. Therefore,

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

and zero otherwise. Thus,

$$E[|X-Y|] = \int_0^L \int_0^L \frac{1}{L^2}|x-y|dxdy$$

First we do the inside integral

$$\begin{aligned}
\int_0^L |x-y|dx &= \int_0^y (y-x)dx + \int_y^L (x-y)dx \\
&= \left(yx - \frac{1}{2}x^2\right)\Big|_{x=0}^{y} + \left(\frac{1}{2}x^2 - yx\right)\Big|_{x=y}^{L} \\
&= \frac{1}{2}y^2 + \frac{1}{2}L^2 - Ly + \frac{1}{2}y^2 \\
&= y^2 + \frac{1}{2}L^2 - Ly
\end{aligned}$$

Thus,

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

We now generalize what we did in chapter 4 by taking $g(x,y) = x + y$. We have

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

Also, suppose that $X \geq Y$. Then, $X - Y \geq 0$. We know that the expected value of a non-negative random variable is non-negative. Therefore, in this case, $E[X-Y] \geq 0$ and so

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

Finally, it is trivial to generalize the above: That is,

$$\boxed{E[a_1X_1 + \ldots + a_nX_n] = a_1E[X_1] + \ldots + a_nE[X_n]}$$

where $a_i \in \mathbb{R}$.

EXAMPLE (The sample mean): Let X_i be independent, identically distributed random variables with expected value μ and distribution function F. We say such a sequence is a *sample* from F. The quantity

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

is called the *sample mean*. We have

$$E[\overline{X}] = E\left[\sum_{i=1}^{n} \frac{1}{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = E[X_i]$$

When μ is unknown, we use the sample mean as an estimate.

EXAMPLE (Boole's inequality): Show that for any subsets $A_1, A_2, \ldots, A_n \subset S$ we have

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

Solution: Define the indicator functions X_i by

$$X_i = 1 \quad \text{if } A_i \text{ occurs}$$

and zero otherwise. Let

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

and so X gives the number of the A_i that have occurred. Finally, let

$$Y = 1 \quad \text{if} \quad X \geq 1$$

and zero otherwise. It is clear that $X \geq Y$. Thus, $E[Y] \leq E[X]$. First note that

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

Also,

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

showing the result.

EXAMPLE: 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. Find the expected number of coupons one need amass before obtaining a complete set of at least one of each type.

EXAMPLE: 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. Find the expected number of coupons one need amass before obtaining a complete set of at least one of each type.

Solution: Let X be the number needed and let $X_i,\ i=0,\ldots,N-1$, be the number of additional coupons that need be obtained after i distinct types have been collected in order to obtain another distinct type. Note that

$$X = X_0 + X_1 + \ldots + X_{N-1}$$

When i distinct types have already been collected, there are $(N-i)$ distinct types left. Therefore, the probability that any choice will yield new type is $(N-i)/N$. Hence, as X_i is a geometric random variable, for $k \geq 1$ we have

$$P\{X_i = k\} = \left(\frac{i}{N}\right)^{k-1} \frac{N-i}{N}$$

Therefore, X_i is geometric with parameter $(N-i)/N$ and so

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

Thus,

$$E[X] = 1 + \frac{N}{N-1} + \frac{N}{N-2} + \ldots + \frac{N}{1} = N\left[1 + \frac{1}{2} + \ldots + \frac{1}{N}\right]$$

EXAMPLE: The sequence of coin flips HHTHHHTTTHTHHHT contains two “runs” of heads of length three, one run of length two and one run of length one. If a coin is flipped 100 times, what is the expected number of runs of length 3?

EXAMPLE: The sequence of coin flips HHTHHHTTTHTHHHT contains two "runs" of heads of length three, one run of length two and one run of length one. If a coin is flipped 100 times, what is the expected number of runs of length 3?

Solution: Let T be the number of runs of length 3. Let I_i be one if a run of length 3 begins at $i \in \{1, \ldots, 98\}$ and zero otherwise. Then, $T = \sum_{i=1}^{98} I_i$ and

$$E[T] = \sum_{i=1}^{98} E[I_i] = \sum_{i=1}^{98} P\{I_i = 1\}$$

A run of length 3 begins at i **only if** there are heads on flips $i,\ i+1,\ i+2$ and tails on $i+3$ (only holds if $i \neq 98$) and $i-1$ (only holds if $i \neq 1$). Thus,

$$E[I_1] = P\{H, H, H, T\} = (1/2)^4$$

$$E[I_{98}] = P\{T, H, H, H\} = (1/2)^4$$

$$E[I_i] = P\{T, H, H, H, T\} = (1/2)^5, \quad i \notin \{1, 98\}$$

Therefore,

$$E[T] = \sum_{i=1}^{98} E[I_i] = 2 \cdot \frac{1}{2^4} + 96 \cdot \frac{1}{2^5} = \frac{25}{8} = 3.125$$

EXAMPLE: Each night for 30 nights an ecologist sets a trap for a rabbit. The trap is very attractive so there is always a rabbit in the trap the next morning. (Only one because the trap isn't large enough for two.) The rabbits that are trapped are marked and released. If there are 100 rabbits in the area and each night the 100 are equally likely to be trapped, what is the expected number of distinct rabbits that are trapped during the 30 night period?

EXAMPLE: Each night for 30 nights an ecologist sets a trap for a rabbit. The trap is very attractive so there is always a rabbit in the trap the next morning. (Only one because the trap isn't large enough for two.) The rabbits that are trapped are marked and released. If there are 100 rabbits in the area and each night the 100 are equally likely to be trapped, what is the expected number of distinct rabbits that are trapped during the 30 night period?

Solution: Let X be the number of distinct rabbits that are trapped during the 30 night period. For $i = 1, \ldots, 100$, let A_i be one if rabbit i is never caught in the 30 nights and zero otherwise. Then $X = 100 - \sum_{i=1}^{100} A_i$ and

$$E[X] = 100 - \sum_{i=1}^{100} E[A_i] = 100 - \sum_{i=1}^{100} P\{A_i = 1\}$$

Now, the probability that a given rabbit is never caught is $(99/100)^{30}$. Therefore,

$$E[X] = 100 - \sum_{i=1}^{100} \left(\frac{99}{100}\right)^{30} = 100 - 100\left(\frac{99}{100}\right)^{30} = 100\left(1 - \left(\frac{99}{100}\right)^{30}\right) \approx 26.0299$$

Section 7.3 Moments of Number of Events

Consider a set of sets A_i. Let X_i be one if A_i occurs. Suppose that $X = \sum_{i=1}^{n} X_i$ is of interest to us.

Now ask how many pairs of events took place? Noting that X_iX_j equals one if and only if both X_i and X_j do, it follows that the number of pairs is

$$\sum_{i<j} X_iX_j$$

Also, because the number of events that do occur is X, the number of pairs of events that occur is exactly

$$\binom{X}{2}$$

Combining these observations yields

$$\binom{X}{2} = \sum_{i<j} X_iX_j$$

where the sum is over the $\binom{n}{2}$ terms. Taking expected values gives us

$$E\left[\binom{X}{2}\right] = \sum_{i<j} E[X_iX_j] = \sum_{i<j} P(A_iA_j)$$

or

$$E\left[\frac{X(X-1)}{2}\right] = \sum_{i<j} P(A_iA_j)$$

which yields

$$E[X^2] - E[X] = 2\sum_{i<j} P(A_iA_j)$$

which gives us $E[X^2]$ and then $\text{Var}(X) = E[X^2] - (E[X])^2$.

EXAMPLE (Moments of a binomial random variable): We have n independent trials each with probability of success p. Let A_i be event of ith success. When $i \neq j$, $P(A_iA_j) = p^2$. Thus,

$$E\left[\frac{X(X-1)}{2}\right] = \sum_{i<j} P(A_iA_j) = \sum_{i<j} p^2 = \frac{n(n-1)}{2}p^2$$

and so

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

Because $E[X] = np$, we have

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

agreeing with the previous calculations.

EXAMPLE: Suppose there are N distinct coupon types and that, independently of past types collected, each new one obtained is type j with probability p_j, obviously with $\sum_{j=1}^{N} p_j = 1$. Find the expected value and variance of the number of different types of coupons that appear among the first n collected.

EXAMPLE: Suppose there are N distinct coupon types and that, independently of past types collected, each new one obtained is type j with probability p_j, obviously with $\sum_{j=1}^{N} p_j = 1$. Find the expected value and variance of the number of different types of coupons that appear among the first n collected.

Solution: Let Y denote the number of different types of collected coupons and let $X = N - Y$ denote the number of uncollected types. Let A_i be the event that the ith coupon is *not* collected, and $X_i = 1$ if A_i occurs. Thus,

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

With probability $(1 - p_i)$, the ith coupon is not collected during each round. Therefore, the probability the ith is not collected in the first n rounds is

$$P(A_i) = P\{X_i = 1\} = (1 - p_i)^n$$

Therefore, $E[X] = \sum_{i=1}^{n} (1 - p_i)^n$, and

$$E[Y] = N - E[X] = N - \sum_{i=1}^{N} (1 - p_i)^n$$

The probability that neither type i nor j is collected in a single round is $(1 - p_i - p_j)$. Therefore,

$$P(A_i A_j) = (1 - p_i - p_j)^n, \;\; i \neq j$$

Therefore,

$$E[X(X-1)] = 2\sum_{i<j} P(A_i A_j) = 2\sum_{i<j} (1 - p_i - p_j)^n$$

and

$$E[X^2] = 2\sum_{i<j} (1 - p_i - p_j)^n + E[X]$$

Hence,

$$\mathrm{Var}(Y) = \mathrm{Var}(X) = E[X^2] - (E[X])^2 = 2\sum_{i<j} (1 - p_i - p_j)^n + \sum_{i=1}^{N} (1 - p_i)^n - \left(\sum_{i=1}^{N} (1 - p_i)^n\right)^2$$

Note that when $p_i = 1/N$ for all i, we get

$$E[Y] = N - N\left(1 - \frac{1}{N}\right)^n = N\left[1 - \left(1 - \frac{1}{N}\right)^n\right]$$

and

$$\mathrm{Var}(Y) = N(N-1)\left(1 - \frac{2}{N}\right)^n + N\left(1 - \frac{1}{N}\right)^n - N^2\left(1 - \frac{1}{N}\right)^{2n}$$

Section 7.4 Covariance, Variance of Sums, and Correlations

THEOREM: If X and Y are independent, then for any functions h and g we have

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

Proof: Let $f(x, y)$ be the joint p.m.f. Then,

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

$$= \int_{-\infty}^{\infty} h(y)f_Y(y)dy \int_{-\infty}^{\infty} g(x)f_X(x)dx = E[h(Y)]E[g(X)]$$

Variance of linear combinations: Suppose that

$$X = X_1 + X_2 + \ldots + X_n$$

We already know that

$$E[X] = \sum_{i=1}^{n} E[X_i]$$

and this *doesn't need independence*. What about variance? Consider $n = 2$. We have

$$\text{Var}(X_1 + X_2) = E[(X - E[X])^2] = E[(X_1 + X_2 - \mu_1 - \mu_2)^2] = E[((X_1 - \mu_1) + (X_2 - \mu_2))^2]$$

$$= E[(X_1 - \mu_1)^2] + E[(X_2 - \mu_2)^2] + 2E[(X_1 - \mu_1)(X_2 - \mu_2)] = \text{Var}(X_1) + \text{Var}(X_2) + 2\text{Cov}(X_1,\ X_2)$$

where the last is a definition. So,

$$\boxed{\text{Var}(X_1 + X_2) = E[(X - E[X])^2] = \text{Var}(X_1) + \text{Var}(X_2) + 2\text{Cov}(X_1,\ X_2)}$$

It is easy to show (see book) that

1. $\text{Cov}(X, Y) = \text{Cov}(Y,\ X)$.
2. $\text{Cov}(X,\ X) = \text{Var}(X)$.
3. $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$.
4. $\text{Cov}\left(\sum_{i=1}^{n} X_i,\ \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \text{Cov}(X_i,\ Y_j)$.

We also have

$$\boxed{\text{Cov}(X_1,\ X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)] = E[X_1X_2] - 2\mu_1\mu_2 + \mu_1\mu_2 = E[X_1X_2] - \mu_1\mu_2}$$

Note that if X_1 and X_2 are independent, then

$$\text{Cov}(X_1,\ X_2) = E[X_1X_2] - \mu_1\mu_2 = E[X_1]E[X_2] - \mu_1\mu_2 = 0$$

EXAMPLE: Two random variables may be dependent, but uncorrelated. Let $R(X) = \{-1, 0, 1\}$ with $P\{X = i\} = 1/3$ for each i. Let $Y = X^2$. Then,

$$\text{Cov}(X, Y) = E[X(Y - E[Y])] = E[X(X^2 - E[X^2])] = E[X^3] - E[X]E[X^2] = 0 - 0 \cdot E[X^2] = 0$$

but not independent, since

$$P\{X = 1,\ Y = 0\} = 0 \neq P\{X = 1\}P\{Y = 0\} = (1/3)(1/3) = 1/9$$

So these are just uncorrelated.

We have

$$\text{Var}(X_1 + X_2 + \ldots + X_n) = \text{Var}(X) = E[(X - E[X])^2]$$

$$= E\left[\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} E[X_i]\right)^2\right] = E\left[\left(\sum_{i=1}^{n}(X_i - \mu_i)\right)^2\right]$$

$$= \sum_{i=1}^{n} \text{Var}(X_i) + \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}(X_i,\ X_j)$$

$$= \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum\sum_{i<j} \text{Cov}(X_i,\ X_j)$$

Now, if the X_i's are *pairwise independent*, then

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

EXAMPLE: Let X be a binomial random variable with parameters n and p, that is, X is the number of successes in n independent trials. Therefore,

$$X = X_1 + \ldots + X_n$$

where X_i is 1 if ith trial was success and zero otherwise. Therefore, X_i's are independent Bernoulli random variables and $E[X_i] = P\{X_i = 1\} = p$. Thus (not from independence),

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np$$

Variance of binomial: Each of the X_i's are independent. Therefore,

$$\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(X_i) = \sum_{i=1}^{n}(p - p^2) = np(1-p)$$

So much easier! In general, we have

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$

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

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)$$

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

Independence implies

$$\text{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i)$$

EXAMPLE: Suppose there are 100 cards numbered 1 through 100. We draw a card at random. Let X be the number of digits ($R(X) = \{1, 2, 3\}$) and Y be the number of zeros ($R(Y) = \{0, 1, 2\}$). The chart giving the probabilities is:

		y			$p_X(x)$
	$p_{XY}(x, y)$	0	1	2	
x	1	9/100	0	0	9/100
	2	81/100	9/100	0	90/100
	3	0	0	1/100	1/100
$p_Y(y)$		90/100	9/100	1/100	

$$E[X] = 1 \cdot \frac{9}{100} + 2 \cdot \frac{90}{100} + 3 \cdot \frac{1}{100} = 1.92$$

$$E[Y] = 0 + 1 \cdot \frac{9}{100} + 2 \cdot \frac{1}{100} = 0.11$$

$$E[XY] = 2 \cdot 1 \cdot \frac{9}{100} + 3 \cdot 2 \cdot \frac{1}{100} = 0.24$$

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0.24 - 1.92 \cdot 0.11 = 0.0288$$

$$(\mathrm{Var}(X) = 0.0936,\ \mathrm{Var}(Y) = 0.1179,\ \rho = \mathrm{Cov}(X, Y)/\sigma_X\sigma_Y = 0.2742)$$

LEMMA: Let $X_1, \ldots, X_n$ be a random sample of size n from a distribution F with mean μ and variance σ^2 (just n experiments). Let $\overline{X} = (X_1 + \ldots + X_n)/n$ be the mean of the random sample. Then

$$E(\overline{X}) = \mu, \quad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

Proof: We have

$$E[\overline{X}] = E\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n} n\mu = \mu$$

and

$$\mathrm{Var}(\overline{X}) = \mathrm{Var}\left(\frac{1}{n}(X_1 + \ldots + X_n)\right) = \frac{1}{n^2}\mathrm{Var}(X_1 + \ldots + X_n) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \quad \blacksquare$$

For real numbers $a,\ b,\ c,\ d$ and random variables X, Y

$$\begin{aligned}\mathrm{Cov}(aX + b,\ cY + d) &= E[(aX + b)(cY + d)] - E[aX + b]E[cY + d] \\ &= E[acXY + bcY + adX + bd] - [aE[X] + b][cE[Y] + d] \\ &= ac(E[XY] - E[X]E[Y]) \\ &= ac\mathrm{Cov}(X, Y)\end{aligned}$$

So

$$\boxed{\mathrm{Cov}(aX + b,\ cY + d) = ac\mathrm{Cov}(X, Y)}$$

Correlation

Suppose that $X_1 = 10X$ and $Y_1 = 5Y$, then

$$\text{Cov}(X_1,\ Y_1) = E[(X_1 - E(X_1))(Y_1 - E(Y_1))] = E[(10X - E(10X))(5Y - E(5Y))] = 50\text{Cov}(X,Y)$$

So knowing if the covariance between X and Y is positive or negative is fine, but doesn't tell you "how correlated." Solution: think in terms of unit-less parameters (standardized random variables). Put

$$X^* = \frac{X - E[X]}{\sigma_X}, \quad Y^* = \frac{Y - E[Y]}{\sigma_Y}$$

then

$$\text{Cov}(X^*,\ Y^*) = \text{Cov}\left(\frac{X - E[X]}{\sigma_X},\ \frac{Y - E[Y]}{\sigma_Y}\right) = \frac{1}{\sigma_X}\frac{1}{\sigma_Y}\text{Cov}(X,Y)$$

Thus, we define the *correlation coefficient* between X and Y as

$$\boxed{\rho = \rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}}$$

So $\text{Cov}(X,Y) > 0$ iff $\rho > 0$ and vice versa.

Properties:

1. $-1 \leq \rho(X,Y) \leq 1$.
2. $\rho(X,Y) = 1$ if and only if $Y = aX + B$ for some $a > 0$.
3. $\rho(X,Y) = -1$ if and only if $Y = aX + B$ for some $a < 0$.

Will just prove item 1:

$$0 \leq \text{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) = \frac{\text{Var}(X)}{\sigma_X^2} + \frac{\text{Var}(Y)}{\sigma_Y^2} + 2\frac{\text{Cov}(X,Y)}{\sigma_x \sigma_Y} = 2[1 + \rho(X,Y)]$$

thus, $\rho(X,Y) \geq -1$. Similarly,

$$0 \leq \text{Var}\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = \frac{\text{Var}(X)}{\sigma_X^2} + \frac{\text{Var}(Y)}{\sigma_Y^2} - 2\frac{\text{Cov}(X,Y)}{\sigma_x \sigma_Y} = 2[1 - \rho(X,Y)]$$

hence $\rho(X,Y) \leq 1$.

EXAMPLE (Investment problem): Suppose someone invests in two financial assets with annual rates of return r_1 and r_2. Let r be the annual rate of return for the overall investment. Let $\sigma^2 = \text{Var}(r)$ and $\sigma_1^2 = \text{Var}(r_1)$ and $\sigma_2^2 = \text{Var}(r_2)$. Let w_1 and w_2 be fractions of investment in assets one and two, that is $r = w_1 r_1 + w_2 r_2$. Then, if $w_1,\ w_2 > 0$ and $\sigma_1,\ \sigma_2 \neq 0$ we have

$$\sigma^2 < \max(\sigma_1^2,\ \sigma_2^2)$$

Proof: Assume that $\sigma_1 \leq \sigma_2$ (if not, go other way). Thus max is σ_2^2. Therefore

$$\begin{aligned}
\sigma^2 = \text{Var}(r) = \text{Var}(w_1 r_1 + w_2 r_2) &= w_1^2 \text{Var}(r_1) + w_2^2 \text{Var}(r_2) + 2 w_1 w_2 \text{Cov}(r_1,\ r_2) \\
&= w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 \rho \sigma_1 \sigma_2 \text{ (now use that } \sigma_1 \leq \sigma_2) \\
&\leq w_1^2 \sigma_2^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 \sigma_2^2 = (w_1 + w_2)^2 \sigma_2^2 = \sigma_2^2 = \max(\sigma_1^2,\ \sigma_2^2) \ \blacksquare
\end{aligned}$$

Section 7.5 Conditional Expectation

Recall that if X and Y are jointly discrete, then the conditional probability mass function of X given $Y = y$ is

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

We therefore define the *conditional expectation* of X given $Y = y$ for all y such that $p_Y(y) > 0$ to be

$$\boxed{E[X|Y = y] = \sum_x xP\{X = x|Y = y\} = \sum_x xp_{X|Y}(x|y)}$$

Therefore, if X and Y are independent, then $E[X|Y = y] = E[X]$.

EXAMPLE: If X and Y are both binomial(n, p) independent random variables (same n and p), calculate the conditional expectation of X given $X + Y = m$.

Solution: We first need the conditional probability mass function of X, given $X + Y = m$. Note that the range k for X is $\le \min(n,\ m)$, since

$$P\{X = k|X + Y = m\} = \frac{P\{X = k, X + Y = m\}}{P\{X + Y = m\}} = \frac{P\{X = k, Y = m - k\}}{P\{X + Y = m\}}$$

We know that the sum of two binomials (with the same p) is binomial. Thus $X + Y$ is binomial$(2n, p)$. So,

$$P\{X + Y = m\} = \binom{2n}{m} p^m (1 - p)^{2n-m}$$

Therefore,

$$P\{X = k|X + Y = m\} = \frac{P\{X = k, Y = m - k\}}{P\{X + Y = m\}} = \frac{P\{X = k\}P\{Y = m - k\}}{P\{X + Y = m\}}$$

$$= \frac{\binom{n}{k} p^k (1 - p)^{n-k} \binom{n}{m - k} p^{m-k} (1 - p)^{n-m+k}}{\binom{2n}{m} p^m (1 - p)^{2n-m}} = \frac{\binom{n}{k}\binom{n}{m - k}}{\binom{2n}{m}}$$

This is a hypergeometic random variable with parameters $m, 2n$, and n. Therefore, the mean is $\dfrac{m \cdot n}{2n} = \dfrac{m}{2}$.

Similarly to the discrete case, the conditional density of X given $Y = y$ is

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

Thus, in this case we define

$$\boxed{E[X|Y = y] = \int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx}$$

provided that $f_Y(y) > 0$. Therefore, if X and Y are independent, then $E[X|Y = y] = E[X]$.

EXAMPLE: Suppose

$$f(x, y) = \frac{1}{y} e^{-x/y} e^{-y}, \quad 0 < x < \infty,\ 0 < y < \infty$$

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

EXAMPLE: Suppose

$$f(x,y) = \frac{1}{y} e^{-x/y} e^{-y}, \quad 0 < x < \infty,\ 0 < y < \infty$$

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

Solution: We have

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

which is exponential with parameter $1/y$. Therefore, the mean is y. That is,

$$E[X|Y=y] = \int_0^{\infty} x \cdot \frac{1}{y} e^{-x/y} dx = y$$

Important: Conditional expectations satisfy all the properties of ordinary expectations. This follows because they *are* just ordinary expectations, as conditional probabilities or densities are themselves probabilities and densities. Therefore,

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

and

$$\boxed{E\left[\sum_i X_i | Y=y\right] = \sum_i E[X_i|Y=y]}$$

Computing Expectations by Conditioning

NOTATION: We will denote by $E[X|Y]$ the function of the random variable Y whose value at $Y=y$ is $E[X|Y=y]$. Note that $E[X|Y]$ is itself a random variable.

THEOREM: We have

$$\boxed{E[X] = E[E[X|Y]]} \tag{1}$$

If Y is discrete, then one can deduce from the above Theorem that

$$\boxed{E[X] = \sum_y E[X|Y=y] P\{Y=y\}} \tag{2}$$

and if Y is continuous with density $f_Y(y)$, then

$$\boxed{E[X] = \int_{-\infty}^{\infty} E[X|Y=y] f_Y(y) dy} \tag{3}$$

Proof (Discrete Case): We have

$$\begin{aligned}
\sum_y E[X|Y=y]P\{Y=y\} &= \sum_y \sum_x xP\{X=x|Y=y\}P\{Y=y\} \\
&= \sum_y \sum_x x\frac{P\{X=x,Y=y\}}{P\{Y=y\}}P\{Y=y\} \\
&= \sum_y \sum_x xP\{X=x,Y=y\} \\
&= \sum_x \sum_y xP\{X=x,Y=y\} \\
&= \sum_x x\sum_y P\{X=x,Y=y\} \\
&= \sum_x xP\{X=x\} \\
&= E[X]
\end{aligned}$$

Proof (Continuous Case): We have

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

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. 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?

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. 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?

Solution: Let X denote the amount of time (in hours) until the miner reaches safety, and let Y denote the door he initially chooses. Now,

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

However,

$$E[X|Y=1] = 3$$

$$E[X|Y=2] = 5 + E[X]$$

$$E[X|Y=3] = 7 + E[X]$$

Therefore

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

hence

$$E[X] = 15$$

EXAMPLE: Consider independent trials, each with a probability p of success. Let N be the time of the first success. Find $\mathrm{Var}(N)$. [Hint: Think about geometric distribution]

EXAMPLE: Consider independent trials, each with a probability p of success. Let N be the time of the first success. Find $\text{Var}(N)$. [Hint: Think about geometric distribution]

Solution: We will condition on the first trial being a success or failure. Let

$$Y = 1 \text{ if the first step is a success}$$

$$Y = 0 \text{ if the first step is a failure}$$

We know that $E[N] = 1/p$ and that

$$\text{Var}(N) = E[N^2] - [E[N]]^2$$

Thus, we need the second moment. We will calculate it by conditioning on Y:

$$E[N^2] = E[E[N^2|Y]]$$

Therefore,

$$E[N^2] = E[N^2|Y=0]P\{Y=0\} + E[N^2|Y=1]P\{Y=1\} = (1-p)E[N^2|Y=0] + pE[N^2|Y=1] \quad (4)$$

Clearly,

$$E[N^2|Y=1] = 1$$

We now show that

$$E[N^2|Y=0] = E[(1+N)^2]$$

In fact, we have

$$E[N^2|Y=0] = \sum_{n=1}^{\infty} n^2 P\{N=n|Y=0\} = \sum_{n=1}^{\infty} n^2 \frac{P\{N=n, Y=0\}}{1-p}$$

Note that $P\{N=n,\ Y=0\} = 0$ if $n=1$ and $(1-p)^{n-1}p$ otherwise. Therefore,

$$E[N^2|Y=0] = \sum_{n=2}^{\infty} n^2(1-p)^{n-2}p = \sum_{n=1}^{\infty}(n+1)^2(1-p)^{n-1}p = E[(N+1)^2]$$

Going back to equation (4), we have

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

Solving this equation for $E[N^2]$, we get

$$E[N^2] = \frac{2-p}{p^2}$$

Finally,

$$\text{Var}(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

EXAMPLE: Let U_i be a sequence of uniform$(0,1)$ independent random variables and let

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

Find $E[N]$.

EXAMPLE: Let U_i be a sequence of uniform$(0,1)$ independent random variables and let

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

Find $E[N]$.

Solution: We will solve a more general problem. For $x \in [0,1]$, let

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

Set

$$m(x) = E[N(x)]$$

We derive an expression by conditioning on U_1. For any x we have

$$m(x) = E[N(x)] = E[E[N(x)|U_1]] = \int_{-\infty}^{\infty} E[N(x)|U_1 = y] f_{U_1}(y)dy = \int_0^1 E[N(x)|U_1 = y]dy$$

Clearly, if $y > x$, then $E[N(x)|U_1 = y] = 1$. What about the other case? One can argue that it should be

$$E[N(x)|U_1 = y] = 1 + m(x-y) \quad \text{if} \quad y \le x$$

In fact, we first note that

$$E[N(x)|U_1 = y] = \sum_{n=2}^{\infty} nP\{N(x) = n|U_1 = y\}$$

Now, assuming $y \le x$, we obtain

$$P\{N(x) = n|U_1 = y\} = P\{U_2 + \ldots + U_n > x - y,\ U_2 + \ldots + U_{n-1} < x - y\} = P\{N(x-y) = n-1\}$$

Thus,

$$E[N(x)|U_1 = y] = \sum_{n=2}^{\infty} nP\{N(x-y) = n-1\} = \sum_{n=1}^{\infty}(n+1)P\{N(x-y) = n\} = m(x-y) + 1$$

therefore

$$m(x) = \int_0^1 E[N(x)|U_1 = y]dy = \int_0^x (1 + m(x-y))dy + \int_x^1 1dy = 1 - x + x + \int_0^x m(x-y)dy$$

which is

$$= 1 + \int_0^x m(u)du \quad (\text{set } u = x - y)$$

Differentiating both sides, we get

$$m'(x) = m(x), \quad m(0) = 1$$

This is equivalent to

$$\frac{m'(x)}{m(x)} = 1 \quad \Longrightarrow \quad \frac{d}{dx}\ln(m(x)) = 1 \quad \Longrightarrow \quad \ln(m(x)) = x + C$$

or

$$m(x) = e^C e^x$$

We use the initial condition $m(0) = 1$ to see that $m(x) = e^x$. Thus, $m(1) = e$.

Calculating Probabilities by Conditioning

Let E be an event and set $X = 1$ if E occurs and 0 if E does not occur. Then

$$E[X] = P(E), \quad E[X|Y = y] = P(E|Y = y)$$

for any Y. Using this in (2) and (3), we get

$$P(E) = \begin{cases} \sum_y P(E|Y = y)P(Y = y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} P(E|Y = y)f_Y(y)dy & \text{if } Y \text{ is continuous} \end{cases} \tag{5}$$

EXAMPLE (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 we've already seen. Suppose that once a prize is rejected, it is lost, 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?

EXAMPLE (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 we've already seen. Suppose that once a prize is rejected, it is lost, 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?

Solution: Fix $k < n$. Strategy: Let the first k prizes go by. Accept the first one if you see that it is larger than all the previous prizes. Let Z_k be the event we choose the best with this strategy. We want to find $P(Z_k)$ and then maximize it with respect to k. We will condition on the position, X, of the best prize. Thus,

$$P(Z_k) = \sum_{i=1}^{n} P(Z_k|X=i)P(X=i) = \frac{1}{n}\sum_{i=1}^{n} P(Z_k|X=i)$$

Note that $P(Z_k|X=i) = 0$ if $i \le k$. Thus,

$$P(Z_k) = \frac{1}{n}\sum_{i=k+1}^{n} P(Z_k|X=i)$$

Now suppose $i > k$. We show that

$$P(Z_k|X=i) = \frac{k}{i-1} \quad \text{if} \quad i > k$$

Indeed, if $X = k+1$, then we *definitely* get the prize and we have $P(Z_k|X=k+1) = \frac{k}{k} = 1$. Similarly, if $X = k+2$, we get

$$P(Z_k|X=k+2) = 1 - \frac{1}{k+1} = \frac{k}{k+1}$$

where $\frac{1}{k+1}$ is representing the probability that the largest of the first $k+1$ was in $k+1$st slot. In general, if $X = i$, then

$$P(Z_k|X=i) = 1 - \frac{i-k-1}{i-1} = \frac{k}{i-1}$$

where $\frac{i-k-1}{i-1}$ is representing the probability that the largest of the first $i-1$ was in the slots $k+1, \ldots, i-1$. It follows that

$$P(Z_k) = \frac{1}{n}\sum_{i=k+1}^{n}\frac{k}{i-1} = \frac{k}{n}\sum_{i=k}^{n-1}\frac{1}{i} = \frac{k}{n}\left(\sum_{i=1}^{n-1}\frac{1}{i} - \sum_{i=1}^{k-1}\frac{1}{i}\right) \approx \frac{k}{n}(\ln(n-1) - \ln(k-1)) \approx \frac{k}{n}\ln\left(\frac{n}{k}\right)$$

Now let $g(x) = \frac{x}{n}\ln\left(\frac{n}{x}\right)$. We want to maximize g with respect to x over the interval $[0, n]$. Differentiating both sides, we get

$$g'(x) = \frac{1}{n}\ln\left(\frac{n}{x}\right) - \frac{1}{n}$$

Therefore,

$$g'(x) = 0 \quad \Longrightarrow \quad \ln\left(\frac{n}{x}\right) = 1 \quad \Longrightarrow \quad x = \frac{n}{e}$$

One can see that $g'(x) > 0$ if $0 < x < n/e$ and $g'(x) < 0$ if $x > n/e$. Therefore, the value found is a maximum. Plugging in n/e to g yields

$$g(n/e) = \frac{1}{e}\ln e = \frac{1}{e}$$

EXAMPLE: Let U be uniform$(0,1)$. Suppose that the conditional distribution of X, given $U = p$, is binomial(n,p). Find the probability mass function of X, i.e. if we flip a coin n times with unknown and uniformly chosen p, what is $P\{X = i\}$ for $i \in \{1, \ldots, n\}$?

Solution: We will condition on U and use the fact that

$$P\{X = i\} = \int_0^1 P\{X = i | U = p\} f_U(p) dp$$

Thus,

$$P\{X = i\} = \int_0^1 P\{X = i | U = p\} f_U(p) dp = \int_0^1 P\{X = i | U = p\} dp = \binom{n}{i} \int_0^1 p^i (1-p)^{n-i} dp$$

It can be shown that

$$\int_0^1 p^i (1-p)^{n-i} dp = \frac{i!(n-i)!}{(n+1)!} = \frac{1}{n+1} \cdot \frac{i!(n-i)!}{n!}$$

Thus,

$$P\{X = i\} = \frac{1}{n+1}$$

That is, we obtain the surprising result that 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 of the values $0, \ldots, n$.

Section 7.7 Moment Generating Functions

DEFINITION: The moment generating function $M(t)$ of the random variable X is defined for all real values of t by

$$\boxed{M(t) = E\left[e^{tX}\right] = \begin{cases} \displaystyle\sum_x e^{tx} p(x) & \text{if } X \text{ is discrete with mass function } p(x) \\ \displaystyle\int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{if } X \text{ is continuous with density } f(x) \end{cases}}$$

We call $M(t)$ the moment generating function because all the moments of X can be obtained by successively differentiating $M(t)$ and then evaluating the result at $t = 0$. For example,

$$M'(t) = \frac{d}{dt} E\left[e^{tX}\right] = E\left[\frac{d}{dt}\left(e^{tX}\right)\right] = E\left[Xe^{tX}\right]$$

where we have assumed that the interchange of the differentiation and expectation operators is legitimate. Hence

$$\boxed{M'(0) = E[X]}$$

Similarly,

$$M''(t) = \frac{d}{dt} M'(t) = \frac{d}{dt} E\left[Xe^{tX}\right] = E\left[\frac{d}{dt}\left(Xe^{tX}\right)\right] = E\left[X^2 e^{tX}\right]$$

Thus,

$$\boxed{M''(0) = E[X^2]}$$

In general, the nth derivative of $M(t)$ is given by

$$M^n(t) = E\left[X^n e^{tX}\right] \quad n \geq 1$$

implying that

$$M^n(0) = E\left[X^n\right] \quad n \geq 1$$

EXAMPLE: If X is a binomial random variable with parameters n and p, then

$$\begin{aligned} M(t) = E\left[e^{tX}\right] &= \sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1-p)^{n-k} \\ &= \sum_{k=0}^{n}\binom{n}{k}(pe^t)^k(1-p)^{n-k} \\ &= (pe^t+1-p)^n \end{aligned}$$

Therefore,

$$M'(t) = n(pe^t+1-p)^{n-1}pe^t$$

Thus,

$$E[X] = M'(0) = np$$

Similarly,

$$M''(t) = n(n-1)(pe^t+1-p)^{n-2}(pe^t)^2 + n(pe^t+1-p)^{n-1}pe^t$$

so

$$E[X^2] = M''(0) = n(n-1)p^2 + np$$

The variance of X is given by

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p)$$

verifying the result obtained previously.

EXAMPLE: If X is a Poisson random variable with parameter λ, then

$$\begin{aligned} M(t) = E\left[e^{tX}\right] &= \sum_{n=0}^{\infty}\frac{e^{tn}e^{-\lambda}\lambda^n}{n!} \\ &= e^{-\lambda}\sum_{n=0}^{\infty}\frac{(\lambda e^t)^n}{n!} \\ &= e^{-\lambda}e^{\lambda e^t} \\ &= \exp\{\lambda(e^t-1)\} \end{aligned}$$

Therefore,

$$M'(t) = \lambda e^t \exp\{\lambda(e^t-1)\}$$

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

Thus,

$$E[X] = M'(0) = \lambda \qquad E[X^2] = M''(0) = \lambda^2 + \lambda$$

so

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

EXAMPLE: If X is an exponential random variable with parameter λ, then

$$M(t) = E\left[e^{tX}\right] = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \lambda\int_0^\infty e^{-(\lambda-t)x}dx = \frac{\lambda}{\lambda-t} \quad \text{for } t<\lambda$$

We note from this derivation that, for the exponential distribution, $M(t)$ defined only for values of t less than λ. Differentiation of $M(t)$ yields

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

Hence,

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

so

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{1}{\lambda^2}$$

EXAMPLE: Let X be a normal random variable with parameters μ and σ^2. We first compute the moment generating function of a standard normal random variable with parameters 0 and 1. Letting Z be such a random variable, we have

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

Hence, the moment generating function of the standard normal random variable Z is $M_Z(t) = e^{t^2/2}$. We now compute the moment generating function of a normal random variable $X = \mu + \sigma Z$ with parameters μ and σ^2. We have

$$M_X(t) = E\left[e^{tX}\right] = E\left[e^{t(\mu+\sigma Z)}\right] = E\left[e^{t\mu}\right]E\left[e^{t\sigma Z}\right] = e^{t\mu}E\left[e^{t\sigma Z}\right] = e^{t\mu}M_Z(t\sigma) = e^{t\mu}e^{(t\sigma)^2/2} = \exp\left\{\frac{\sigma^2t^2}{2}+\mu t\right\}$$

By differentiating, we obtain

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

Thus,

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

so

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

An important property of moment generating functions is that the moment generating function of the sum of independent random variables equals the product of the individual moment generating functions. To prove this, suppose that X and Y are independent and have moment generating functions $M_X(t)$ and $M_Y(t)$, respectively. Then $M_{X+Y}(t)$, the moment generating function of $X+Y$, is given by

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

Another important result is that the moment generating function uniquely determines the distribution. That is, if $M_X(t)$ exists and is finite in some region about $t=0$, then the distribution of X is uniquely determined.

EXAMPLE: Let X and Y be independent binomial random variables with parameters (n,p) and (m,p), respectively. What is the distribution of $X+Y$?

Solution: The moment generating function of $X+Y$ is given by

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

However, $(pe^t+1-p)^{m+n}$ is the moment generating function of a binomial random variable having parameters $m+n$ and p. Thus, this must be the distribution of $X+Y$.

EXAMPLE: Let X and Y be independent Poisson random variables with means respective λ_1 and λ_2. What is the distribution of $X+Y$?

Solution: The moment generating function of $X+Y$ is given by

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

Hence, $X+Y$ is Poisson distributed with mean $\lambda_1+\lambda_2$.

EXAMPLE: Let X and Y be independent normal random variables with respective parameters (μ_1,σ_1^2) and (μ_2,σ_2^2). What is the distribution of $X+Y$?

Solution: The moment generating function of $X+Y$ is given by

$$M_{X+Y}(t) = M_X(t)M_Y(t) = \exp\left\{\frac{\sigma_1^2t^2}{2}+\mu_1t\right\}\exp\left\{\frac{\sigma_2^2t^2}{2}+\mu_2t\right\} = \exp\left\{\frac{(\sigma_1^2+\sigma_2^2)t^2}{2}+(\mu_1+\mu_2)t\right\}$$

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

Discrete Probability Distribution

	Probability mass function, $p(x)$	Moment generating function, $M(t)$	Mean	Variance
Binomial with parameters n, p; $0 \leq p \leq 1$	$\binom{n}{x} p^x (1-p)^{n-x}$ $x = 0, 1, \ldots, n$	$(pe^t + 1 - p)^n$	np	$np(1-p)$
Poisson with parameter $\lambda > 0$	$e^{-\lambda} \frac{\lambda^x}{x!}$ $x = 0, 1, 2, \ldots$	$\exp\{\lambda(e^t - 1)\}$	λ	λ
Geometric with parameter $0 \leq p \leq 1$	$p(1-p)^{x-1}$ $x = 1, 2, \ldots$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative binomial with parameters r, p; $0 \leq p \leq 1$	$\binom{n-1}{r-1} p^r (1-p)^{n-r}$ $n = r, r+1, \ldots$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

Continuous Probability Distribution

	Probability mass function, $p(x)$	Moment generating function, $M(t)$	Mean	Variance
Uniform over (a, b)	$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential with parameter $\lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$	$\frac{\lambda}{\lambda - t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma with parameters $(s, \lambda), \lambda > 0$	$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{s-1}}{\Gamma(s)} & x \geq 0 \\ 0 & x < 0 \end{cases}$	$\left(\frac{\lambda}{\lambda - t}\right)^s$	$\frac{s}{\lambda}$	$\frac{s}{\lambda^2}$
Normal with parameters (μ, σ^2)	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$ $-\infty < x < \infty$	$\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$	μ	σ^2