PROPERTIES OF EXPECTATIONS

1. Conditional expectations

1.1. **Discrete case.** Recall that if X and Y are discrete random variables then the conditional probability mass function of X given that $Y = y$ is defined as

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

for all values of y such that $p_Y(y) > 0$. Since X given $Y = y$ is also a random variable, we define the conditional expectation of X given $Y = y$ as

$$\mathbb{E}[X|Y = y] = \sum_x x \cdot p_{X|Y}(x|y).$$

Thus, $\mathbb{E}[X|Y = y]$ is a function of y. We define $\mathbb{E}[X|Y]$ as a function of random variable Y that take values $\mathbb{E}[X|Y = y]$ with probability $\mathbb{P}(Y = y)$.

Note that by our definition, the conditional expectation $\mathbb{E}[X|Y]$ of X given Y is itself a random variable. For example, its expectation is

$$\begin{aligned}
\mathbb{E}\Big[\mathbb{E}[X|Y]\Big] &= \sum_y \mathbb{E}[X|Y = y] \cdot \mathbb{P}(Y = y) \\
&= \sum_y \left(\sum_x x \cdot p_{X|Y}(x|y)\right) \cdot p_Y(y) \\
&= \sum_x x \cdot \left(\sum_y p_{X|Y}(x|y) p_Y(y)\right) \\
&= \sum_x x \cdot \left(\sum_y p(x, y)\right) \\
&= \sum_x x \cdot p_X(x) \\
&= \mathbb{E}X.
\end{aligned}$$

We have just proved the **law of total expectation**

$$\mathbb{E}\Big[\mathbb{E}[X|Y]\Big] = \mathbb{E}[X].$$

Example 1.1. Suppose that the number of people entering a department store on a given day is a random variable with mean 50. Suppose further that the amounts of money spent by these customers are independent random variables having a common mean of 8 dollars. Finally, suppose also that the amount of money spent by a customer is also independent of the total number of customers who enter the store. What is the expected amount of money spent in the store on a given day?

Solution. Let N denote the number of customers that enter the store and X_i the amount spent by the i-th such customer. The total amount of money spent can be expressed as $\sum_{i=1}^N X_i$. Note that we can not directly apply the formula that the expectation of a sum of random variables is the sum of their expectations here because the number of summands must be nonrandom for the formula to hold.

By law of total expectation,

$$\mathbb{E}\left[\sum_{i=1}^N X_i\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^N X_i \Big| N\right]\right].$$

We first calculate the conditional expectation. For any integer n, we have

$$\mathbb{E}\left[\sum_{i=1}^N X_i \Big| N = n\right] = \mathbb{E}\left[\sum_{i=1}^n X_i \Big| N = n\right] = \sum_{i=1}^n \mathbb{E}\left[X_i \Big| N = n\right] = 8n,$$

which implies

$$\mathbb{E}\left[\sum_{i=1}^N X_i \Big| N\right] = 8N.$$

Therefore

$$\mathbb{E}\left[\sum_{i=1}^N X_i\right] = \mathbb{E}[8N] = 8 \cdot 50 = 400.$$

□

1.2. **Continuous case.** Recall that if X and Y have a joint probability density function $f(x, y)$, then the conditional probability density function of X given that $Y = y$ is defined, for all values of y such that $f_Y(y) > 0$, by

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

We then define the conditional expectation of X given $Y = y$ as

$$\mathbb{E}[X|Y = y] = \int x \cdot f_{X|Y}(x|y) dx.$$

Similar to the case of discrete random variables, we define $\mathbb{E}[X|Y]$ as a function of random variable Y, i.e. $\mathbb{E}[X|Y] = g(Y)$ where the function is $g(y) = \mathbb{E}[X|Y = y]$. The law of total probability still holds

$$\mathbb{E}\Big[\mathbb{E}[X|Y]\Big] = \mathbb{E}[X].$$

Example 1.2. Suppose that the joint density of X and Y is given by

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

Find $\mathbb{E}[X|Y]$.

Solution. We first find the marginal density of Y by integrating the joint density with respect to x:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^{\infty} \frac{e^{-x/y-y}}{y}dx = e^{-y}, \quad y \in (0,\infty).$$

Therefore the conditional density of X given $Y = y \in (0,\infty)$ is

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{\frac{e^{-x/y-y}}{y}}{e^{-y}} = \frac{e^{-x/y}}{y}, \quad x \in (0,\infty).$$

Hence

$$\mathbb{E}[X|Y=y] = \int_0^{\infty} \frac{e^{-x/y}}{y}dx = -e^{-x/y}\Big|_0^{\infty} = 1.$$

Thus, $\mathbb{E}[X|Y] = 1$. $\square$

2. Conditional variance

Just as we have defined the conditional expectation of X given the value of Y, we can also define the conditional variance of X given Y

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

In other words, $Var(X|Y)$ is exactly analogous to the usual definition of variance, but now all expectations are conditional on the fact that Y is known. It is easy to check that

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

Note that $Var(X|Y)$ is a random variable. In particular, we can calculate its expectation

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

Also, since $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$, we have

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

Adding the last two equations, we obtain the **law of total variance**

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

Example 2.1. Suppose that by any time t the number of people that have arrived at a train depot is a Poisson random variable with mean λt. If the initial train arrives at the depot at a time (independent of when the passengers arrive) that is uniformly distributed over $(0,T)$, what are the mean and variance of the number of passengers who enter the train?

Solution. For each $t \geq 0$, let $N(t)$ denote the number of arrivals by t, and let Y denote the time at which the train arrives. The random variable of interest is then $N(Y)$. Conditioning on Y gives

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

Hence $\mathbb{E}[N(Y)|Y] = \lambda Y$, so taking the expectation gives

$$\mathbb{E}[N(Y)] = \mathbb{E}[\lambda Y] = \frac{\lambda T}{2}.$$

To obtain $Var(N(Y))$, we use the law of total variance by conditioning on Y. First

$$\begin{aligned} Var(N(Y)|Y=t) &= Var(N(t)|Y=t) \\ &= Var(N(t)), \quad \text{by the independence of } Y \text{ and } N(t) \\ &= \lambda t. \end{aligned}$$

This implies $Var(N(Y)|Y) = \lambda Y$. Therefore by the law of total variance,

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

□

3. Covariance and correlation

The covariance of X and Y is defined as

$$Cov(X,Y) = \mathbb{E}\Big[\big(X - \mathbb{E}X\big)\big(Y - \mathbb{E}Y\big)\Big].$$

By expanding the expression in the brackets and using properties of expectation, it is easy to see that

$$Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}X \cdot \mathbb{E}Y.$$

If X and Y are independent then $\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y$, therefore $Cov(X,Y) = 0$.

We list some properties of covariance, the proofs of which can be found in the textbook.

Proposition 3.1.
(a) $Cov(X,Y) = Cov(Y,X)$.
(b) $Cov(X,X) = Var(X)$.
(c) $Cov(aX,Y) = aCov(X,Y)$.
(d) $Cov\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m Cov(X_i, Y_j)$.

Example 3.2. Let $X_1, ..., X_n$ be independent and identically distributed random variables having expected value μ and variance σ^2. Define the sample mean and sample variance by

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

Find $Var(\overline{X})$ and $\mathbb{E}S^2$.

Solution. Since $X_1, ..., X_n$ are independent,

$$Var(\overline{X}) = \frac{1}{n^2} Var\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n Var(X_i) = \frac{\sigma^2}{n}.$$

To find $\mathbb{E}S^2$, we start with the following identity

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

Taking the expectations of both sides, we get

$$\begin{aligned}
\mathbb{E}\sum_{i=1}^{n}(X_i-\overline{X})^2 &= \sum_{i=1}^{n}\mathbb{E}(X_i-\mu)^2 - n\mathbb{E}(\overline{X}-\mu)^2 \\
&= n\sigma^2 - nVar(\overline{X}) \\
&= (n-1)\sigma^2.
\end{aligned}$$

It then follows that $\mathbb{E}S^2 = \sigma^2$. □

4. Moment generating function

The moment generating function $M(t)$ of the random variable X is defined for all real values of t by $M(t) = \mathbb{E}e^{tX}$. If X is discrete with probability mass function p then

$$M(t) = \sum_x e^{tx}p(x),$$

and if X is continuous with density f then

$$M(t) = \int e^{tx}f(x)dx.$$

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

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

and at $t = 0$ we get $M'(0) = \mathbb{E}[X]$. In general, the nth derivative of $M(t)$ is given by a very nice formula

$$M^{(n)}(t) = \mathbb{E}[X^n e^{tX}].$$

At $t = 0$ we get

$$M^{(n)}(0) = \mathbb{E}X^n,$$

which is known as the n-th moment of X.

The distribution of a random variable is uniquely determined by its moment generating function. That is, X and Y have the same distribution if and only if their moment generating functions are identical $M_X(t) = M_Y(t)$ for all t.

Example 4.1. If X is a binomial random variable with parameters n and p, then by the binomial theorem,

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

Taking the derivative of M we have

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

Hence the first moment of X is $\mathbb{E}X = M'(0) = np$. Taking the derivative one more time,

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

so

$$\mathbb{E}X^2 = M''(0) = n(n-1)p^2 + np.$$

We can then calculate the variance of X:

$$Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = n(n-1)p^2 + np - (np)^2 = np(1-p).$$

Example 4.2. If X is a Poisson random variable with mean λ then

$$M(t) = \sum_{n=0}^{\infty} e^{tn} \cdot \frac{e^{-\lambda}\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} = e^{-\lambda} \cdot \exp(\lambda e^t) = \exp(\lambda(e^t - 1)).$$

By differentiating $M(t)$ and then evaluating the derivatives at $t = 0$, we can check that $\mathbb{E}X = Var(X) = \lambda$.

Example 4.3. The moment generating function for exponential distribution with parameter λ is

$$M(t) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{(t-\lambda)x} dx = \frac{\lambda}{\lambda - t}$$

for $t < \lambda$. Note that $M(t) = +\infty$ if $t \geq \lambda$ because the integral diverges in that case. However, we can still calculate the the moments of X because we only need to evaluate the derivatives of $M(t)$ at zero, and for that purpose, it is sufficient that $M(t)$ exists for all t near zero (for example for all $t < \lambda$).

Example 4.4. Let Z be a standard normal random variable. Then

$$\begin{aligned} M_Z(t) &= \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-t)^2}{2} + \frac{t^2}{2}\right) dx \\ &= e^{t^2/2} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-t)^2}{2}\right) dx \\ &= e^{t^2/2} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2}\right) du \quad \text{using } u = x - t \\ &= e^{t^2/2}. \end{aligned}$$

Now, let $X = \sigma Z + \mu \sim N(\mu, \sigma^2)$ be a normal random variable with mean μ and variance σ^2, then

$$M_X(t) = \mathbb{E}e^{t(\sigma Z + \mu)} = e^{\mu t} \cdot \mathbb{E}e^{(t\sigma)Z} = e^{\mu t} \cdot M_Z(t\sigma) = e^{\mu t + \sigma^2 t^2/2}.$$

Sum of independent random variables. 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. Indeed, if X and Y are independent then

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

Example 4.5. Let X and Y be Poisson random variables with mean λ_X and λ_y. Then from Example **??** we have

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = \exp(\lambda_X(e^t - 1)) \cdot \exp(\lambda_Y(e^t - 1)) = \exp((\lambda_X + \lambda_Y)(e^t - 1)).$$

Since $M_{X+Y}(t)$ uniquely determines the distribution of $X+Y$ and $M_{X+Y}(t)$ has the form of the moment generating function of a Poisson random variable with parameter $\lambda_X + \lambda_Y$, we conclude that $X+Y$ is a Poisson random variable with parameter $\lambda_X + \lambda_Y$.