PROBABILITY

QSNAKE EDITION

0. Property of Expectation.

Since X is a weighted average of the possible values of X, it follows that if X must lie between a and b, then so must its expected value. That is, if

$$P\{a \leq X \leq b\} = 1$$

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

Proof. Suppose X is a discrete random variable for which $P\{a \leq X \leq b\} = 1$. Since this implies that $p(x) = 0$ for all x outside of the interval $[a, b]$, it follows that

$$\begin{aligned}
E[x] &= \sum_{x:p(x)>0} xp(x) \\
&\geq \sum_{x:p(x)>0} ap(x) \\
&= a \sum_{x:p(x)>0} p(x) \\
&= a
\end{aligned}$$

and the proof in the continuous case is similar, the result follows. ■

1. Expectation of Sums of Random Variables.

Recall. X is a non-negative random variable, we have

$$E[X] = \int_0^\infty P\{X > t\} dt$$

Proof.

$$\begin{aligned}
E[X] &= \int_0^\infty xf(x)dx \\
&= \int_0^\infty \left(\int_0^x 1dt\right) f(x)dx \\
&= \int_0^\infty \int_0^x f(x)dtdx \\
&= \int_0^\infty \left(\int_t^\infty f(x)dx\right) dt \\
&= \int_0^\infty P\{X > t\}dt
\end{aligned}$$

■

Proposition. *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

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

Proof. Use the recall above we get

$$E[g(X,Y)] = \int_0^\infty P\{g(X,Y) > t\}dt$$

Writing

$$P\{g(X,Y) > t\} = \iint_{(x,y):g(x,y)>t} f(x,y)dydx$$

shows that

$$E[g(X,Y)] = \int_0^\infty \iint_{(x,y):g(x,y)>t} f(x,y)dydx$$

Interchanging the order of integration gives

$$
\begin{aligned}
E[g(X,Y)] &= \int_x \int_y \int_{t=0}^{g(x,y)} f(x,y) dt dy dx \\
&= \int_x \int_y g(x,y) f(x,y) dy dx
\end{aligned}
$$

■

Corollary. *$E[X]$ and $E[Y]$ are finite,*

$$E[X+Y] = E[X] + E[Y]$$

Proof. Suppose $E[X]$ and $E[Y]$ are both finite and let $g(X,Y) = X+Y$. Then, in the continuous case,

$$
\begin{aligned}
E[X+Y] &= \int_{\infty}^{\infty} \int_{\infty}^{\infty} (x+y) f(x,y) dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y) dy dx + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x,y) dx dy \\
&= \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy \\
&= E[X] + E[Y]
\end{aligned}
$$

■

2*. Moments of the Number of Events that Occur.

For given events $A_1, \cdots, A_n$, find $E[X]$, where X is the number of these events that occur. The solution then involved defining an indicator variable I_i for event A_i such that

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

Because

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

we obtained the result

$$E[X] = E\left[\sum_{i=1}^{n} I_i\right] = \sum_{i=1}^{n} E[I_i] = \sum_{i=1}^{n} 1 \cdot P(A_i) + 0 \cdot P(A_i^c) = \sum_{i=1}^{n} P(A_i)$$

Consider the number of pairs of events that occur. Because $I_i I_j$ will equal 1 if both A_i and A_j occur and will equal 0 otherwise, it follows that the number of pairs is equal to $\sum_{i<j} I_i I_j$. But because X is the number of events that occur, it also follows that the number of pairs of events that occur is $(\genfrac{}{}{0pt}{}{X}{2})$. Consequently,

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

and taking expectations yields

$$E\left[\binom{X}{2}\right] = \sum_{i<j} E[I_i I_j] = \sum_{i<j} P(A_i A_j)$$

or

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

giving that

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

Moreover, by considering the number of distinct subsets of k events that all occur, we see that

$$\binom{X}{k} = \sum_{i_1 < i_2 < \cdots < i_k} I_{i_1} I_{i_2} \cdots I_{i_k}$$

Taking expectations gives the identity

$$E\left[\binom{X}{k}\right] = \sum_{i_1 < i_2 < \cdots < i_k} E[I_{i_1} I_{i_2} \cdots I_{i_k}] = \sum_{i_1 < i_2 < \cdots < i_k} P A_{i_1} A_{i_2} \cdots A_{i_k}$$

3. Covariance, Variance of Sums, and Correlations.

Proposition. *If X and Y are independent, then for any functions h and g,*

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

Proof. Suppose that X and Y are jointly continuous with joint density $f(x,y)$. Then

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

The proof in the discrete case is similar. ■

<u>Definition</u>. *The covariance between X and Y, denote by $Cov(X,Y)$, is defined by*

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])]$$

Upon expanding the right side of the preceding definition, we see that

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

Remark. *If X and Y are independent, then $Cov(X,Y) = 0$. However, the converse is not true.*

Example.

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

and defining

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

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

Proposition.

(i) $Cov(X, Y) = Cov(Y, X)$
(ii) $Cov(X, X) = Var(X)$
(iii) $Cov(aX, Y = aCov(X, Y)$
(iv) $Cov\Big(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m}\Big) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov\ (X_i, Y_j)$

Proof. We only proof (iv) here, let $\mu_i = E[X_i]$ and $v_j = E[Y_j]$. Then

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mu_i, \quad E\left[\sum_{j=1}^{m} Y_i\right] = \sum_{j=1}^{m} v_j$$

and

$$\begin{aligned} Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) &= E\left[\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right)\left(\sum_{j=1}^{m} Y_j - \sum_{j=1}^{m} v_j\right)\right] \\ &= E\left[\sum_{i=1}^{n}(X_i - \mu_i)\sum_{j=1}^{m}(Y_j - v_j)\right] \\ &= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m}(X_i - \mu_i)(Y_j - v_j)\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} E[(X_i - \mu_i)(Y_j - v_j)] \end{aligned}$$

■

Corollary.

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

Proof.

$$\begin{aligned} Var\left(\sum_{i=1}^{n} X_i\right) &= Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j) \\ &= \sum_{i=1}^{n} Var(X_i) + \sum\sum_{i \neq j} Cov(X_i, X_j) \end{aligned}$$

■

Corollary. *If X_1 and X_2 are random variables with joint pdf $f(x_1, x_2)$, then*

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$$

4. Conditional Expectation.

Definition. *If X and Y are jointly discrete random variables, then the conditional probability mass function of X, given that $Y = y$, is defined for all y such that $P\{Y = y\} > 0$, by*

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

It is therefore natural to define, in this case, the conditional expectation of X given that $Y = y$, for all value of y such that $p_Y(y) > 0$, by

$$\begin{aligned} E[X|Y = y] &= \sum_x xP\{X = x \mid Y = y\} \\ &= \sum_x xp_{X|Y}(x|y) \end{aligned}$$

Proposition.

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

Proof. We only proof the equation when X and Y are discrete
Observe that $E[X|Y]$ is a random variable of Y, so we can rewrite

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

Now, the equation can be written as

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

■

Definition (Conditional Variance).

$$Var(X \mid Y) = E[(X - E[X|Y])^2|Y]$$

Proposition.

$$Var(X) = E[Var(X|Y)] + Var(E[X|Y])$$

Proof.

(1) Observe $E[Var(X|Y)]$ first, rewrite

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

so

$$\begin{aligned} E[Var(X|Y)] &= E[E[X^2|Y]] - E[(E[X|Y])^2] \\ &= E[X^2] - E[(E[X|Y])^2] \quad (\star_1) \end{aligned}$$

(2) Observe $Var(E[X|Y])$

$$\begin{aligned} Var(E[X|Y]) &= E[(E(X|Y))^2] - (E[E[X|Y]])^2 \\ &= E[(E[X|Y])^2] - (E[X])^2 \quad (\star_2) \end{aligned}$$

Combine $(\star_1)$ and $(\star_2)$, we get

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

■

Theorem. *If X and Y are jointly distributed random variables, and $g(x)$ is a function, then*

$$E[g(X)Y \mid x] = g(x)E(Y \mid x)$$

5. Moment Generating Function.

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

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

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.

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

where we have assumed that the interchange of the differentiation and expectation operator is legitimate. That is, we have assumed that

$$\frac{d}{dt}\left[\sum_x e^{tx} p(x)\right] = \sum_x \frac{d}{dt}[e^{tx} p(x)]$$

in the discrete case and

$$\frac{d}{dt}\left[\int e^{tx} f(x)dx\right] = \int \frac{d}{dt}[e^{tx} f(x)]dx$$

in the continuous case. This assumption can almost always be justified. Hence, evaluated at $t = 0$, we obtain

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

Similarly,

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

Thus,

$$M''(x) = E[X^2]$$

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

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

implying that

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

Definition. *For any n random variables $X_1, \cdots, X_n$ the joint moment generating function, $M(t_1, \cdots, t_n)$, is defined, for all real values of $t_1, \cdots, t_n$, by*

$$M(t_1, \cdots, t_n) = E[e^{t_1X_1 + \cdots + t_nX_n}]$$

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

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

where the t is in the ith place.

Proposition. *the joint moment generating function $M(t_1, \cdots, t_n)$ uniquely determines the joint distribution of $X_1, \cdots, X_n$. and the n random variables $X_1, \cdots, X_n$ are independent if and only if*

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

6*. General Definition of Expectation. There also exist random variables that are neither discrete nor continuous, and they too, may possess an expectation.

Example. X be a Bernoulli random variable with parameter $p = \frac{1}{2}$, and let Y be a uniformly distributed random variable over the interval $[0, 1]$. and suppose that X and Y are independent, and define the new random variable W by

$$W = \begin{cases} X & \text{if } X = 1 \\ Y & \text{if } X \neq 1 \end{cases}$$

Clearly, W is neither a discrete (since its set of possible values, [0,1], is uncountable) nor a continuous (since $P\{W = 1\} = \frac{1}{2}$)

In order to define the expectation of an arbitrary random variable, we require the notion of a Stieltjes integral. For instance, for any function g, $\int_a^b g(x)dx$ is defined by

$$\int_a^b g(x)dx = \lim \sum_{i=1}^{n} g(x_i)(x_i - x_{i-1})$$

where the limit is taken over all $a = x_0 < x_1 < x_2 < \cdots < x_n = b$ as $n \to \infty$ and where $\max_{i=1,\cdots,n}(x_i - x_{i-1} \to 0)$.

For any distribution function F, we define the Stieltjes integral of the nonnegative function g over the interval $[a, b]$ by

$$\int_a^b g(x)dF(x) = \lim \sum_{i=1}^{n} g(x_i)[F(x_i) - F(x_{i-1})]$$