I. Joint Distribution Example

$$p(1,1) = 0.3 \quad p(2,1) = 0.1$$
$$p(1,2) = 0.1 \quad p(2,2) = 0.5$$

Expectation of $g(x, y) = xy$:

$$E[XY] = 1 \times 0.3 + 2 \times 0.1 + 2 \times 0.1 + 4 \times 0.5 = 2.7$$

Marginal probability mass functions:

$$\begin{aligned} p_X(1) &= p(1,1) + p(1,2) = 0.4 \\ p_X(2) &= p(2,1) + p(2,2) = 0.6 \\ p_Y(1) &= p(1,1) + p(2,1) = 0.4 \\ p_Y(2) &= p(1,2) + p(2,2) = 0.6 \end{aligned}$$

II. Conditional distributions

A. **Definition:** The *conditional probability mass function* for discrete r.v.'s X given $Y = y$ is

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

for all y such that $P(Y = y) > 0$.

B. **Example:** Using the same joint pmf given above, compute $p_{X|Y}(x|y)$.

$$\begin{aligned} p_{X|Y}(1|1) &= \frac{p(1,1)}{p_Y(1)} = \frac{0.3}{0.4} = 0.75 \\ p_{X|Y}(1|2) &= \frac{p(1,2)}{p_Y(2)} = \frac{0.1}{0.6} \approx 0.167 \\ p_{X|Y}(2|1) &= \frac{p(2,1)}{p_X(1)} = \frac{0.1}{0.4} = 0.25 \\ p_{X|Y}(2|2) &= \frac{p(2,2)}{p_X(2)} = \frac{0.5}{0.6} \approx 0.83. \end{aligned}$$

C. **Example:** Suppose $Y \sim \text{Bin}(n_1, p)$ and $X \sim \text{Bin}(n_2, p)$ and let $Z = X+Y$ and $q = 1-p$. Calculate $p_{X|Z}(x|z)$.

$$\begin{aligned} p_{X|Z}(x|z) &= \frac{p_{X,Z}(x, z)}{p_Z(z)} \\ &= \frac{p_{X,Y}(x, z - x)}{p_Z(z)} \end{aligned}$$

$$
\begin{aligned}
&= \frac{p_X(x)p_Y(z-x)}{p_Z(z)} \\
&= \frac{\binom{n_1}{x}p^x q^{n_1-x}\binom{n_2}{z-x}p^{z-x}q^{n_2-z+x}}{\binom{n_1+n_2}{z}p^z q^{n_1+n_2-z}} \\
&= \frac{\binom{n_1}{x}\binom{n_2}{z-y}}{\binom{n_1+n_2}{m}}.
\end{aligned}
$$

The last formula is the pmf of the *Hypergeometric distribution*, the distribution that is canonically associated with the following experiment. Suppose you draw m balls from an urn containing n_1 black balls and n_2 red balls. The Hypergeometric is the distribution of the random number counting the number of black balls you selected.

D. **Definition:** The *conditional probability density function* for continuous r.v.'s X given $Y = y$ is

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

E. **Example:**

$$
f(x,y) = \begin{cases} 6xy(2-x-y) & 0 < x < 1, 0 < y < 1 \\ 0 & \text{ow.} \end{cases}
$$

$$
f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{6xy(2-x-y)}{\int_0^1 6xy(2-x-y)dx} = \frac{6x(2-x-y)}{4-3y}.
$$

III. Conditional expectation

A. **Definition:** The *conditional expectation* of discrete r.v. X given $Y = y$ is

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

B. **Example:** Using the same joint pmf given above, compute $E[X|Y=1]$.

$$
E[X|Y=1] = 1 \times p_{X|Y}(1|1) + 2 \times p_{X|Y}(2|1) = 1 \times 0.75 + 2 \times 0.25 = 1.25.
$$

C. **Definition:** The *conditional expectation* of continuous r.v.'s X given $Y = y$ is

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

D. **Example:** Using the same joint pdf as in the previous section, find $E[X|Y=y]$.

$$
E[X|Y=y] = \int_0^1 x f_{X|Y}(x|y)dx = \int_0^1 \frac{6x^2(2-x-y)}{4-3y}dx = \frac{5-4y}{8-6y}.
$$

E. **Note:** We can think of $E[X|Y]$, where a value for the r.v. Y is not specified, as a random function of the r.v. Y.

F. **Proposition:**

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

Proof:

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

G. **Example:** Suppose that the expected number of accidents per week is 4 and suppose that the number of workers injured per accident has mean 2. Also assume that the number of people injured is independently determined for each accident. What is the expected number of injuries per week?

HOMEWORK

H. **Definition:** a *compound random variable* is a sum of random length of independent and identically distributed random variables.

I. **Property:** If $X = \sum_{i=1}^{N} X_i$ is a compound random variable and $E[X_i] = \mu$, then

$$E[X] = \mu E[N].$$

J. **Example:** Suppose you are in a room in a cave. There are three exits. The first takes you on a path that takes 3 hours but eventually deposits you back in the same room. The second takes you on a path that takes 5 hours but also deposits you back in the room. The third takes you to the exit, which is 2 hours away. On average, how long do you expect to stay in the cave if you randomly select an exit each time you enter the room in the cave?

HOMEWORK
Hint: It is helpful to condition on the random variable Y indicating which exit you choose on your first attempt to exit the room.

K. **Proposition:** There is an equivalent results for variances.

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

L. **Result:** We can use the above result to find the variance of a compound r.v. Suppose that $E[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$.

$$\text{Var}(X|N) = \text{Var}\left(\sum_{i=1}^{N} X_i|N\right) = \sum_{i=1}^{N} \text{Var}(X_i) = N\sigma^2.$$

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

and you obtain the variance of the compound random variable X by knowing the mean and variance of X_i and N.

IV. **Comprehensive Example:**
You are presented n prizes in random sequence. When presented with a prize, you are told its rank relative to all the prizes you have seen (or you can observe this, for example if the prizes are money). If you accept the prize, the game is over. If you reject the prize, you are presented with the next prize. How can you best improve your chance of getting the best prize of all n.
Let's attempt a strategy where you reject the first k prizes. Thereafter, you accept the first prize that is better than the first k. We will find the k that maximizes your chance of getting the best prize and compute your probability of getting the best prize for that k.
Let $P_k(\text{best})$ be the probability that the best prize is selected using the above strategy.

HOMEWORK
Hint: Think LTP. It is easiest to compute $P_k(\text{best})$ by conditioning on the position i of the best prize X. Also note that the probability of choosing the best prize is only achieved if the best of the first $i-1$ is among the first k. Finally, $\sum_{i=k}^{n} \frac{1}{i} \approx \int_k^n \frac{1}{x} dx$.

The answer without elaboration: Using the proposed strategy you can achieve nearly 40% chance of getting the best prize out of all n prizes.