**Handout 6: Joint and Conditional Distributions**

**Math/Stat 394: Probability I**

Wellner; 2/9/2000

**Joint Probability Distributions, Discrete Case:** Suppose that $X, Y$ are two discrete random variables defined on the same sample space. Then the *joint probability mass function* or simply *joint probability distribution* $p_{X,Y}$ of $(X, Y)$ is given by

$$p_{X,Y}(x,y) = P(X = x, Y = y) = P([X = x] \cap [Y = y]) = P(X = x \text{ and } Y = y)$$

for possible values $(x, y)$ of $(X, Y)$.
The *marginal probability mass functions* $p_X$ and $p_Y$ of $X$ and $Y$ respectively, are obtained from the joint mass function by summing out:

$$p_X(x) = \sum_y p_{X,Y}(x,y) \qquad \text{and} \qquad p_Y(y) = \sum_x p_{X,Y}(x,y)\,.$$

**Example 1.** Suppose that $X_1$ and $X_2$ are the numbers which appear on two successive (independent) rolls of one fair die. Then the joint probability mass function of $(X_1, X_2)$ is

$$p_{X_1,X_2}(x_1,x_2) = P(X_1 = x_1, X_2 = x_2) = \frac{1}{36} \quad \text{for} \quad x_i \in \{1,\ldots,y\}, i = 1,2\,.$$

Note that this mass function has equal heights of 1/36 at all points of the 6x6 lattice; it's a kind of "spiked carpet" with all the spikes of equal height. Also note that $X_1$ and $X_2$ are independent random variables:

$$p_{X_1,X_2}(x_1,x_2) = \frac{1}{36} = p_{X_1}(x_1)p_{X_2}(x_2)$$

for all $(x_1, x_2)$ with $x_i \in \{1, \ldots, y\}, i = 1, 2$.

**Example 2.** In the set-up of Example 1, define $T = X_1 + X_2$ and $M = \max\{X_1, X_2\}$. If we record the values of $(T, M)$ for each possible outcome, we get the following table:

| $x_2/x_1$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 6 | (7,6) | (8,6) | (9,6) | (10,6) | (11,6) | (12,6) |
| 5 | (6,5) | (7,5) | (8,5) | (9,5) | (10,5) | (11,6) |
| 4 | (5,4) | (6,4) | (7,4) | (8,4) | (9,4) | (10,5) |
| 3 | (4,3) | (5,3) | (6,3) | (7,4) | (8,5) | (9,6) |
| 2 | (3,2) | (4,2) | (5,3) | (6,4) | (7,5) | (8,6) |
| 1 | (2,1) | (3,2) | (4,3) | (5,4) | (6,5) | (7,6) |

This yields the following joint and marginal distributions for $(T, M)$: the entries in the table are $36 \cdot p_{T,M}(t, m)$, $36 \cdot p_M(m)$, and $36 \cdot p_T(t)$.

| $m/t$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | $p_M(m)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | | | | | | 2 | 2 | 2 | 2 | 2 | 1 | 11 |
| 5 | | | | | 2 | 2 | 2 | 2 | 1 | | | 9 |
| 4 | | | | 2 | 2 | 2 | 1 | | | | | 7 |
| 3 | | | 2 | 2 | 1 | | | | | | | 5 |
| 2 | | 2 | 1 | | | | | | | | | 3 |
| 1 | 1 | | | | | | | | | | | 1 |
| $p_T(t)$ | 1 | 2 | 3 | 4 | 5 | 6 | 5 | 4 | 3 | 2 | 1 | 36 |

**Conditional Probabilities:** For events $A, B$, recall that

$$P(A|B) \equiv P(AB)/P(B) \qquad \text{so that} \qquad P(AB) = P(A|B)P(B)$$

for events $B$ with $P(B) > 0$. Recall that *independence* of two events $A, B$ means that

$$P(A|B) = P(A) \qquad \text{or} \qquad P(AB) = P(A)P(B)\,.$$

Let $(X, Y)$ have discrete joint probability mass function

$$p_{X,Y}(x, y) \equiv P(X = x, Y = y)$$

with marginals $p_X(x) = P(X = x)$ and $p_Y(y) = P(Y = y)$. We define

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

and

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

to be the *conditional distributions*. The *conditional mean (or regression) function(s)* $m(x) = E(Y|X = x)$ and $E(X|Y = y)$ are given by

$$m(x) = E(Y|X = x) = \sum_y y p_{Y|X}(y|x),$$

and

$$E(X|Y = y) = \sum_x x p_{X|Y}(x|y)\,.$$

**Example 3.** For the joint distribution of $(T, M)$ in Example 2, We easily compute

$$P(M = m|T = 7) = \begin{cases} \frac{2/36}{6/36} = 1/3, & m \in \{4, 5, 6\} \\ 0, & \text{otherwise} \end{cases},$$

and

$$P(M = m|T = 6) = \begin{cases} \frac{2/36}{5/36} = 2/5, & m \in \{4, 5\} \\ \frac{1/36}{5/36} = 1/5, & m = 3 \\ 0, & \text{otherwise} \end{cases}.$$

Conditioning on $M$ rather than $T$, we find, for example, that

$$P(T = t|M = 6) = \begin{cases} \frac{2/36}{11/36} = 2/11, & t \in \{7, \ldots, 11\} \\ \frac{1/36}{11/36} = 1/11, & t = 12 \\ 0, & \text{otherwise} \end{cases},$$

and

$$P(T = t|M = 3) = \begin{cases} \frac{2/36}{5/36} = 2/5, & t \in \{4, 5\} \\ \frac{1/36}{11/36} = 1/5, & t = 6 \\ 0, & \text{otherwise} \end{cases}.$$

Note that

$$E(T|M = 6) \;=\; (7 + 8 + 9 + 10 + 11) \cdot \frac{2}{11} + 12 \cdot \frac{1}{11} = 9.27$$

$$\begin{aligned}
E(T|M=5) &= (6+7+8+9)\cdot\frac{2}{9}+10\cdot\frac{1}{9}=7.78\,,\\
E(T|M=4) &= (5+6+7)\cdot\frac{2}{7}+8\frac{1}{7}=6.29\,,\\
E(T|M=3) &= (4+5)\cdot\frac{2}{5}+6\frac{1}{5}=4.8\,,\\
E(T|M=2) &= 3\cdot\frac{2}{3}+4\frac{1}{3}=3.33\,.\\
E(T|M=1) &= 2\cdot 1=2\,.
\end{aligned}$$

In the other direction we compute

$$\begin{aligned}
E(M|T=2) &= 1\cdot 1=1\,,\\
E(M|T=3) &= 2\cdot(1/3)+3\cdot(2/3)=8/3\doteq 2.67\,,\\
E(M|T=4) &= (3+4)\cdot(1/2)=3.5\,,\\
E(M|T=5) &= 3\cdot(1/5)+(4+5)\cdot(2/5)=4.2\,,\\
E(M|T=6) &= (4+5+6)\cdot(1/3)=5\,,\\
E(M|T=8) &= 4\cdot(1/5)+(5+6)\cdot(2/5)=5.2\,.\\
E(M|T=9) &= (5+6)\cdot(1/2)=5.5\,,\\
E(M|T=10) &= 5\cdot(1/3)+6\cdot(2/3)=17/3\doteq 5.67\,,\\
E(M|T=11) &= 6\,,\\
E(M|T=12) &= 6\,.
\end{aligned}$$

These represent a practical and easy to use aspect of conditional distributions. These are numerical in character and very valuable, but not at all cute or pleasing.

We now turn to theoretically pleasing results that are very usable.

**Example 1.** Let $X \sim Poisson(\lambda)$ and $Y \sim Poisson(\mu)$ be independent. Let $T \equiv X+Y$. We know already that $T = X+Y \sim Poisson(\lambda+\mu)$. Also $(X|T=n) \sim \text{Binomial}(n, p=\lambda/(\lambda+\mu))$.

*Proof:*

$$\begin{aligned}
P(X=k|T=n) &= \frac{P(X=k,T=n)}{P(T=n)}=\frac{P(X=k,Y=n-k)}{P(T=n)}\\
&= \frac{P(X=k)P(Y=n-k)}{P(T=n)}\\
&= \frac{e^{-\lambda}\lambda^k/k!e^{-\mu}\mu^{n-k}/(n-k)!}{e^{-(\lambda+\mu)}(\lambda+\mu)^n/n!}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{n!}{k!(n-k)!}\lambda^k \mu^{n-k}/(\lambda+\mu)^n \\
&= \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{with } \; p \equiv \lambda/(\lambda+\mu)\,.
\end{aligned}
$$

**Example 2.** Let $U$ ~Binomial$(M, p)$ and $V$ ~Binomial$(N, p)$ be independent. Let $W \equiv U + V$. We know already that $W$ ~Binomial$(M + N, p)$. Also $(U|W = n)$ ~Hypergeometric$(R = M, R + W = M + N, n)$.
*Proof:*

$$
\begin{aligned}
P(U = k|W = n) &= \frac{P(U = k, W = n)}{P(W = n)} = \frac{P(U = k, V = n-k)}{P(W = n)} \\
&= \frac{P(U = k)P(V = n-k)}{P(W = n)} \\
&= \frac{\binom{M}{k} p^k q^{M-k} \binom{N}{n-k} p^{n-k} q^{N-(n-k)}}{\binom{M+N}{n} p^n q^{M+N-n}} \\
&= \frac{\binom{M}{k}\binom{N}{n-k}}{\binom{M+N}{n}}\,.
\end{aligned}
$$

**Example 3.** We study disease $D$ for a year. It has a high incidence rate.

Let $X =\#$ of diseased men, and $Y =\#$ of diseased women. Let $T = X + Y$.
**Question 1:** Is this a sex-linked disease, in that men are more susceptible? If not, then $X|T = n$ is Binomial$(n, 1/2)$ (from Example 1). Suppose we observed $X = x = 63$ diseased men among 100 diseased people. Now

$$
P(X \geq 63|T = 100) = P(Bin(100, 1/2) \geq 63) \doteq P(N(50, 5^2) \geq 62.5) = .0062\,.
$$

This is good evidence that the disease $D$ is sex-linked. The number .0062 is called the $p-$*value*. (It answers: how unlikely was the result we observed?)

This disease is "manageable", until death occurs, let us suppose.
**Question 2:** Are *deaths* more likely among diseased men than among diseased women?
We had $R = 63$ diseased (red) men and $W = 37$ diseased (white) women in the total of $N = 100$ people. The total numer of deaths (within one year of diagnosis) was 10, say. That is $W$ too on a value of $w = n = 10$. Now if $U$ denotes the number of male deaths, then from Example 2,

$$
(U|W = n) \sim \text{Hypergeometric}(R = 63, N = 100, n = 10)\,,
$$

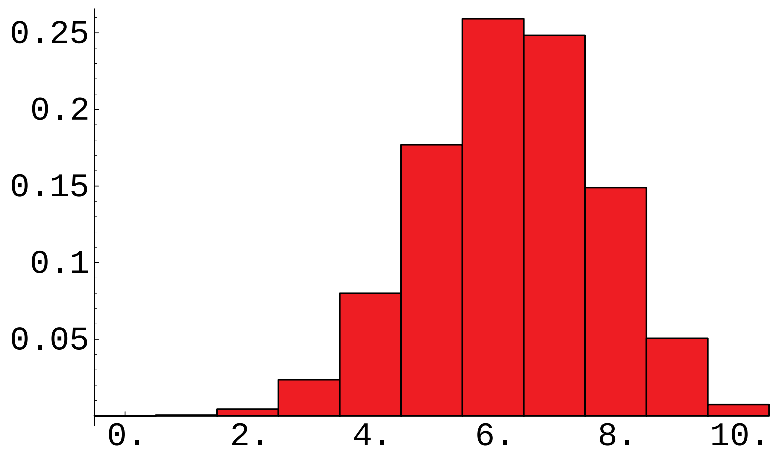


Figure 1: Plot of Hypergeometric(63, 100, 10) distribution.

and the corresponding mass function is

$$p_{U|W}(k|10) = \frac{\binom{63}{k}\binom{37}{10-k}}{\binom{100}{10}} \qquad \text{for} \qquad k = 0, \ldots, 10\,,$$

which is shown in Figure 1.

This probability is $= .007383$ if $k = 10$, and $= .050589$ if $k = 9$; for a p-value of .0074 if $U = 10$ is observed, and a p-value of .0580 if $U = 9$ is observed. This is good evidence if $U = 10$, and some evidence if $U = 9$. relative to male deaths being more likely.

**The Multiple Hypergeometric and Multinomial Distributions:** Suppose that an urn contains 3 types of balls: $R$ red balls, $W$ white balls, and $B$ black balls. Let $N = R + W + B$ be the total number of balls in the urn. Suppose we sample $n$ balls from the urn *without replacement*. Let $X$ be the number of red balls in the sample, let $Y$ be the number of white balls in the sample, and let $Z$ be the number of black balls in the sample. Then the joint probability mass function of $X, Y, Z$ is the *multiple hypergeometric distribution* given as follows:

$$p(x, y, z) \equiv P(X = x, Y = y, Z = z) = \frac{\binom{R}{x}\binom{W}{y}\binom{B}{z}}{\binom{N}{n}} \qquad \text{provided } x+y+z = n\,.$$

Since we must always have $z = n - x - y$ and $Z = n - X - Y$, we may equivalently write

$$p(x, y) \equiv P(X = x, Y = y, Z = n - x - y) = \frac{\binom{R}{x}\binom{W}{y}\binom{B}{n-x-y}}{\binom{N}{n}}$$
$$\text{provided } x \leq R, y \leq W, 0 \leq n - x - y \leq B, \,.$$

**Example 1.** Suppose $R = 2$, $W = 4$, and $B = 6$, so $N = 12$. Suppose we sample $n = 3$ balls. Then

$$p(x, y) = \frac{\binom{2}{x}\binom{4}{y}\binom{6}{3-x-y}}{\binom{12}{3}}, \qquad \text{provided } x \in \{0, 1, 2\}, y \in \{0, 2, 3\}, x+y \in \{0, 1, 2, 3\}\,.$$

Thus

$$p(1, 2) = P(X = 1, Y = 2) = \frac{\binom{2}{1}\binom{4}{2}\binom{6}{0}}{\binom{12}{3}} = \frac{12}{220}\,.$$

Note that the marginal distributions of $X$ and $Y$ are Hypergeometric($R = 2, N = 10, n = 3$) and Hypergeometric($R = 4, N = 10, n = 3$) respectively: Thus we have

$$p_X(x) = \frac{120}{220}, \frac{90}{220}, \frac{10}{220}, 0, \quad \text{for } x = 0, 1, 2, 3\,,$$

while

$$p_Y(y) = \frac{56}{220}, \frac{112}{220}, \frac{48}{220}, \frac{4}{220} \quad \text{for } x = 0, 1, 2, 3\,,$$

Here is a table of the joint distribution of $X, Y$:

| x/y | 0 | 1 | 2 | 3 | |
|---|---|---|---|---|---|
| 0 | 20/220 | 60/220 | 36/220 | 4/220 | 120/220 |
| 1 | 30/220 | 48/220 | 12/220 | 0 | 90/220 |
| 2 | 6/220 | 4/220 | 0 | 0 | 10/220 |
| 3 | 0 | 0 | 0 | 0 | 0 |
| | 56/220 | 112/220 | 48/229 | 4/220 | 1 |

Let $T \equiv X + Y$. Then

$$p_T(2) = p(0,2) + p(1,1) + p(2,0) = \frac{90}{220}\,.$$

Alternatively, we can revisualize the urn as containing 6 red or white balls and 6 black balls, so that $T \sim$Hypergeometric$(R = 6, N = 12, n = 3)$, and therefore

$$p_T(t) = P(T = t) = \frac{\binom{6}{t}\binom{6}{3-t}}{\binom{12}{3}} \quad \text{for} \quad t \in \{0, 1, 2, 3\}\,.$$

Here is a table of the resulting distribution $p_T$ of $T$:

| $t$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $p_T(t)$ | 20/220 | 90/220 | 90/220 | 20/220 |

Note that $P(X < Y) = (60 + 36 + 4 + 12 + 0 + 0)/220 = 112/220$, or, in general

$$P((X,Y) \in C) = \sum_{(x,y) \in C} p_{X,Y}(x,y)\,.$$

Now suppose that we sample *with replacement* from the urn. Let $p_1 = R/N$, $p_2 = W/N$, and $p_3 = B/N$ be the fractions of the red, white, and black balls respectively; note that $p_1 + p_2 + p_3 = (R + W + B)/N = N/N = 1$. If we let $X \equiv$ the number of red balls in the sample, $Y \equiv$ the number of white balls in the sample, and $Z \equiv$ the number of black balls in the sample, then $X + Y + Z = n$, and the joint mass function of $(X, Y, Z)$ is the $Multinomial_3(n, \underline{p} = (p_1, p_2, p_3))$ distribution given by:

$$p(x,y,z) = P(X = x, Y = y, Z = z) = \binom{n}{x,\, y,\, z} p_1^x p_2^y p_3^z \qquad \text{provided } x+y+z = n\,.$$

Equivalently,

$$\begin{aligned} p(x,y) &= P(X = x, Y = y, Z = n - x - y) \\ &= \binom{n}{x,\, y,\, (n - x - y)} p_1^x p_2^y p_3^{n-x-y} \\ &\qquad x \in \{0, \ldots, n\}, y \in \{0, \ldots, n\}, \text{ and } x + y \in \{0, \ldots, n\}\,. \end{aligned}$$

Here

$$\binom{n}{x,\, y,\, z} \equiv \frac{n!}{x!y!z!}\,.$$

**Example 2.** In particular, suppose that the urn is as above. Then $p_1 = 1/6$, $p_2 = 2/6$, and $p_3 = 3/6 = 1/2$. If we sample $n = 8$ times from the urn (with replacement), then $(X, Y, Z) \sim Multinomial_3(8, (1/6, 2/6, 3/6))$, and, for example,

$$p(1,2) = P(X = 1, Y = 2) = \binom{8}{1,2,5}(1/6)^1(2/6)^2(3/6)^5 = .0972\,.$$

Note that $X \sim Binomial(n, p_1)$; similarly, $Y \sim Binomial(n, p_2)$ and $Z \sim Binomial(n, p_3)$. That is, the marginals of a multinomial distribution are binomial distributions.