4 Multiple random variables

4.1 Discrete multivariate distributions

Definition. Let $\mathcal{S}$ be a sample space. A **random vector** $(X_1, X_2, \ldots, X_k)$ is a k-dimensional function defined on $\mathcal{S}$ (or a "vector of random variables all defined simultaneously on $\mathcal{S}$").

Example. Consider SWR from a population of size N. Possible responses: yes/no/unknown with proportions p_1, p_2, p_3 ($p_1 + p_2 + p_3 = 1$). There are N^n possible samples (of size n). For $s \in \mathcal{S}$ let

$X_1(s) =$ "# yes in the sample",
$X_2(s) =$ "# no in the sample",
$X_3(s) =$ "# unknowns in the sample".

Then (X_1, X_2, X_3) is a trivariate random vector with $X_1 + X_2 + X_3 = n$. Select x_1 "yes", x_2 "no" and x_3 "unknown". There are $(Np_1)^{x_1}(Np_2)^{x_2}(Np_3)^{x_3}$ ways to select these and $n!/(x_1!x_2!x_3!)$ ways to order a selection. This gives the **trinomial** pmf

$$\begin{aligned} P(X_1 = x_1, X_2 = x_2, X_3 = x_3) &= \frac{1}{N^n}\frac{n!}{x_1!x_2!x_3!}(Np_1)^{x_1}(Np_2)^{x_2}(Np_3)^{x_3} \\ &= \frac{1}{N^n}\frac{n!}{x_1!x_2!x_3!}N^{\overbrace{x_1 + x_2 + x_3}^{n}}p_1^{x_1}p_2^{x_2}p_3^{x_3} \\ &= \frac{n!}{x_1!x_2!x_3!}p_1^{x_1}p_2^{x_2}p_3^{x_3}. \end{aligned}$$

Definition. Let $(X_1, X_2, \ldots, X_k)$ be a random vector with each X_i discrete.
(i) The joint pmf for $(X_1, \ldots, X_k)$ is $f(x_1, \ldots, x_k) = P(X_1 = x_1, \ldots, X_k = x_k)$.
(ii) The (marginal) pmf for each X_i is $f_{X_i}(x) = P(X_i = x)$ (in line with the one-dimensional case).

Theorem. Suppose (X, Y) has joint pmf $f_{X,Y}$. Then $f_X(x) = \sum_y f_{X,Y}(x, y)$ and $f_Y(y) = \sum_x f_{X,Y}(x, y)$ ("sum out").

Proof: partitions rule.

Definition. Suppose (X, Y) has joint pmf $f_{X,Y}(x, y)$. The **conditional pmf** of X given $Y = y$ is

$$f_{X|Y}(x|y) = P(X = x|Y = y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{if } f_Y(y) > 0, \\ f_X(x) & \text{if } f_Y(y) = 0. \end{cases}$$

Theorem. Suppose X and Y have a joint discrete distribution. For fixed y, the conditional pmf $f_{X|Y}(x|y)$ is a valid pmf in x.

Proof: "≥ 0" holds by definition. By using the previous definition and theorem we obtain

$$\begin{aligned} f_Y(y) > 0: &\qquad \sum_x f_{X|Y}(x|y) = \frac{\sum_x f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1, \\ f_Y(y) = 0: &\qquad \sum_x f_{X|Y}(x|y) = \sum_x f_X(x) = 1. \qquad \square \end{aligned}$$

Definition. R.v.'s $X_1, \ldots, X_k$ are independent if, for all subsets $A_1, \ldots, A_k$,

$$P(X_1 \in A_1, \ldots, X_k \in A_k) = P(X_1 \in A_1) \cdots P(X_k \in A_k).$$

"Any k events about $X_1, \ldots, X_k$ are independent."

Example. Roll two fair dice with $X_i =$ "result of dice i", $i = 1, 2$.
(i) X_1 and X_2 ("the two dice") are independent: consider arbitrary A, B.
Then $P(X_1 = j, X_2 = k) = 1/36$, $P(X_1 = j) = P(X_2 = j) = 1/6$
$(j, k = 1, \ldots, 6)$, and

$$\begin{aligned} P(X_1 \in A, X_2 \in B) &= \sum_{j\in A}\sum_{k\in B} P(X_1 = j, X_2 = k) \quad (= \frac{1}{36} = \frac{1}{6}\cdot\frac{1}{6}) \\ &= \sum_{j\in A}\sum_{k\in B} P(X_1 = j)P(X_2 = k) = \sum_{j\in A} P(X_1 = j) \sum_{k\in B} P(X_2 = k) \\ &= P(X_1 \in A)P(X_2 \in B). \end{aligned}$$

(ii) X_1 and $Y = X_1 + X_2$ are not independent:
$P(X_1 = 6, Y = 2) = 0 \neq P(X_1 = 6)P(Y = 2) > 0.$

Theorem. Suppose (X, Y) has joint pmf $f_{X,Y}(x, y)$. Then
(i) X, Y are independent $\iff f_{X,Y}(x, y) = f_X(x)f_Y(y)$ for all x, y,
(ii) X, Y are independent $\iff f_{X|Y}(x|y) = f_X(x)$ for all x,
(iii) X, Y are independent $\iff f_{X,Y}(x, y) = g(x)h(y)$ for some functions g, h.

Proof: (i) "$\Rightarrow$" holds by definition. "$\Leftarrow$":

$$\begin{aligned} P(X \in A, Y \in B) &= \sum_{x\in A}\sum_{y\in B} f_{X,Y}(x, y) = \sum_{x\in A} f_X(x) \sum_{y\in B} f_Y(y) \\ &= P(X \in A)P(Y \in B). \end{aligned}$$

(ii) exercise.
(iii) "$\Rightarrow$" holds by (i). "$\Leftarrow$":

$$\begin{aligned} f_X(x)f_Y(y) &= \sum_y f_{X,Y}(x, y) \sum_x f_{X,Y}(x, y) = \sum_y g(x)h(y) \sum_x g(x)h(y) \\ &= g(x)h(y) \sum_y h(y) \sum_x g(x) = f_{X,Y}(x, y) \sum_y \sum_x g(x)h(y) \\ &= f_{X,Y}(x, y) \underbrace{\sum\sum f_{X,Y}(x, y)}_{1}. \quad \square \end{aligned}$$

Example. Roll two fair dice with results (X_1, X_2) and total Y. Obtain the pmf of Y by using the joint pmf of (X_1, X_2). *Solution:*

$$\begin{aligned} f_Y(y) &= P(X_1 + X_2 = y) = \sum\sum\nolimits_{x_1+x_2=y} f_{X_1,X_2}(x_1, x_2) \\ &= \sum_{x_1} \sum_{x_2=y-x_1} f_{X_1,X_2}(x_1, x_2) = \sum_{x_1} f_{X_1,X_2}(x_1, y - x_1) \\ &= \sum_{x_1} f_{X_1}(x_1)f_{X_2}(y - x_1) = \sum_{x_1=1}^{6} \frac{1}{6} 1_{\{1,\ldots,6\}}(x_1) \frac{1}{6} 1_{\{1,\ldots,6\}}(y - x_1) \\ &= 1_{\{2,\ldots,12\}}(y) \frac{1}{36}(6 - |7 - y|). \end{aligned}$$

(In the last step just count all possibilities for which $1 \le x_1 \le 6$ and $1 \le y - x_1 \le 6$). The arguments just used yield the following theorem:

Theorem. Consider (X, Y) with joint pmf $f_{X,Y}(x, y)$ and $T = X + Y$.
(i) The pmf for T is $f_T(t) = \sum_x f_{X,Y}(x, t-x) = \sum_y f_{X,Y}(t-y, y)$.
(ii) (**Convolution formula.**) If X and Y are independent, then

$$f_T(t) = \sum_x f_X(x) f_Y(t-x) = \sum_y f_X(t-y) f_Y(y).$$

Example / Theorem. Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$, X, Y independent. Then $X + Y \sim \text{Poisson}(\lambda + \mu)$. ("$X + Y$ stays in the family", with $E(X+Y) = EX + EY = \lambda + \mu$.)

Proof. Let $T = X + Y$. Then

$$\begin{aligned}
f_T(t) &= \sum_x f_X(x) f_Y(t-x) = \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} \frac{\mu^{t-x} e^{-\mu}}{(t-x)!} 1_{\{0,1,\ldots\}}(t-x) \\
&= \frac{(\lambda+\mu)^t e^{-(\lambda+\mu)}}{t!} \sum_{x=0}^{t} \frac{t!}{(\lambda+\mu)^t} \frac{\lambda^x \mu^{t-x}}{x!(t-x)!} \\
&= \frac{(\lambda+\mu)^t e^{-(\lambda+\mu)}}{t!} \sum_{x=0}^{t} \binom{t}{x} \Big(\underbrace{\frac{\mu}{\lambda+\mu}}_{1-p}\Big)^{t-x} \Big(\underbrace{\frac{\lambda}{\lambda+\mu}}_{p}\Big)^{x}
\end{aligned}$$

which is the desired formula since the sum (of $\text{bin}(t, p)$ probabilities) equals 1.

Note. The above method can be used for arbitrary functions of discrete r.v.'s X, Y:

$$f_{g(X,Y)}(w) = P\{g(X, Y) = w\} = \sum_{g(x,y)=w} f_{X,Y}(x, y).$$

Sometimes it is easier to compute the cdf:

Example. Consider X, Y independent, $W = \min(X, Y)$, $X \sim \text{geometric}(p)$, $Y \sim \text{geometric}(q)$. We know $P(X > n) = (1-p)^n$, $P(Y > m) = (1-q)^m$. Therefore

$$\begin{aligned}
P(W \le w) &= 1 - P(W > w) = 1 - P(X > w, Y > w) \\
&= 1 - P(X > w) P(Y > w) = 1 - \{(1-p)(1-q)\}^w.
\end{aligned}$$

Setting $\tilde{p} = (1-p)(1-q)$ gives $P(W > w) = \tilde{p}^w = \{1 - (1-\tilde{p})\}^w$, that is $W \sim \text{geometric}(1-\tilde{p}) = \text{geometric}\{1 - (1-p)(1-q)\}$.

4.2 Continuous multivariate distributions

Definition. The random vector $(X_1, X_2, \ldots, X_k)$ has an absolutely continuous distribution with (joint) pdf $f(x_1, \ldots, x_k)$ if for all $a_i \le b_i$, $1 \le i \le k$,

$$P(a_1 < X_1 < b_1, \ldots, a_k < X_k < b_k) = \int_{a_k}^{b_k} \cdots \int_{a_1}^{b_1} f(x_1, \ldots, x_k)\, dx_1 \ldots x_k.$$

In this case, for 'any nice' $A \subset \mathbb{R}^k$:

$$P\{(X_1, \ldots, X_k) \in A\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} 1_A(x_1, \ldots, x_k) f(x_1, \ldots, x_k)\, dx_1 \ldots x_k.$$

Q: How to obtain a *marginal* pdf?
A: *Integrate out* the other variables.

Theorem. Consider (X, Y) with joint pdf $f_{X,Y}(x, y)$. Then X and Y are each (absolutely) continous with **marginals**

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx.$$

(Analogously for k dimensions.)

Proof. Note that $\int_a^b f(x)\, dx = \int_a^b g(x)\, dx$ for all $[a, b]$ implies $f(x) = g(x)$ (almost surely).

$$\begin{aligned}\int_a^b f(x)\, dx &= P(a \le X \le b) = P(a \le X \le b, -\infty \le Y \le \infty) \\ &= \int_a^b \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy}_{\hat{=}\, g(x)}\, dx. \qquad \square\end{aligned}$$

Example. Consider (X, Y) with pdf $f_{X,Y}(x, y) = \frac{1}{2}(\lambda^2 e^{-\lambda(x+y)} + \mu^2 e^{-\mu(x+y)})$ $1_{(0,\infty)}(x)1_{(0,\infty)}(y)$. Then

$$\begin{aligned}f_X(x) &= \int_0^{\infty} \frac{1}{2}(\lambda^2 e^{-\lambda(x+y)} + \mu^2 e^{-\mu(x+y)})1_{(0,\infty)}(y)\, dy\, 1_{(0,\infty)}(x) \\ &= \frac{1}{2}(-\lambda e^{-\lambda(x+y)} - \mu e^{-\mu(x+y)})\Big|_{y=0}^{\infty} 1_{(0,\infty)}(x) \\ &= \frac{1}{2}(\lambda e^{-\lambda x} + \mu e^{-\mu x})1_{(0,\infty)}(x) \;=\; f_Y(x)\end{aligned}$$

(average of two exponentials). These are valid pdf's since $\int_0^{\infty} f_X(x) = 1 = \int_0^{\infty}\int_0^{\infty} f_{X,Y}(x, y)$.

Definition. Suppose (X, Y) has joint pdf $f_{X,Y}(x, y)$. The **conditional pdf** of X given $Y = y$ is

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{if} \quad f_Y(y) > 0 \\ f_X(x) & \text{if} \quad f_Y(y) = 0 \end{cases}$$

The **conditional distribution** of X given $Y = y$ is given by probabilities such as

$$P(a \le X \le b \mid Y = y) = \int_a^b f_{X|Y}(x|y)\, dx$$

where $f_{X|Y}(x|y)$ is a valid pdf (as in the discrete case).

Problem (with the left-hand side):
the expression is technically not defined ($P(Y = y) = 0$).

Explanation: Let $\delta > 0$ be small, suppose $f_{X,Y}$ is continuous. Then

$$P(a \le X \le b \mid y \le Y \le y + \delta) = \frac{\int_y^{y+\delta}\int_a^b f_{X,Y}(x, u)\, dx\, du}{\int_y^{y+\delta} f_Y(u)\, du}$$