Elements of Probability and Statistics

Lecture 11: Joint Distributions

IAI, TCG-CREST September 30, 2023

For distributions defined on more than one variables, the joint, marginal, and conditional distributions are defined and discussed, first for the discrete case, and then for the continuous and hybrid cases.

11.1 Joint Discrete Distributions

Joint discrete distributions can be specified in terms of their CDFs or their PMFs. Here they are described for 2 variables, and the definitions can be extended for n variables.

Defn. Joint CDF: The joint CDF $F_{X,Y}$ of two random variables X and Y is the function,

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y).$$

Defn. Joint PMF: The joint PMF $p_{X,Y}$ of two random variables X and Y is the function,

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

Valid PMFs must be non-negative and must sum to 1.

The probability of any arbitrary event represented by set A in the support of (X, Y) can be calculated as,

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

From a joint distribution over X and Y, the distribution of X alone can be obtained by summing over (or marginalizing out) the possible supports of Y.

Defn. Marginal Distribution: The Marginal PMF of random variable X is,

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

Marginal CDFs can be computed by taking a limit:

$$F_X(x) = P(X \leq x) = \lim_{y\to\infty} P(X \leq x, Y \leq y) = \lim_{y\to\infty} F_{X,Y}(x, y).$$

However marginal PMFs are usually easier to work with.

If Y is observed, we may want to update the distribution of X according to the observed value of Y. This is described by the conditional PMF.

Defn. Conditional PMF: The conditional PMF of X given $Y = y$ is,

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

Prop. Conditional PMFs are related by Bayes Rule:

$$P(Y = y|X = x) = \frac{P(X = x|Y = y)P(Y = y)}{P(X = x)}.$$

The Law of Total Probability relates marginal PMFs to the conditional PMFs:

$$P(X = x) = \sum_y P(X = x|Y = y)P(Y = y).$$

Ex. The following is a concrete example of a joint distribution over Bernoulli random variables X and Y. The four possible values of the joint PMF are described in a 2×2 contingency table.

	$Y = 0$	$Y = 1$	Marg. X
$X = 0$	5/100	20/100	25/100
$X = 1$	3/100	72/100	75/100
Marg. Y	8/100	92/100	

Summing over the rows provides the marginal distribution of X, and similarly summing over the columns provides the marginal distribution of Y. Any conditional PMF can be easily calculated by dividing the joint PMF entry with the corresponding marginal PMF value.

Defn. Independence of discrete random variables: Random variables X and Y are independent if the joint CDF factors into a product of the marginal CDFs,

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)\ \forall x, y.$$

Independence of discrete random variables can also be defined by the condition of the joint PMF factoring into the product of marginal PMFs,

$$P(X = x, Y = y) = P(X = x)P(Y = y)\ \forall x, y,$$

By an equivalent definition of independence, all conditional PMFs are equal to the marginal PMF,

$$P(X = x|Y = y) = P(X = x)\ \forall x, y,\ \text{where } P(X = x) > 0.$$

An interesting relationship between Binomial and Poisson distributions can be established.

Thm. If $N \sim \text{Pois}(\lambda)$ and $X|N = n \sim \text{Bin}(n, p)$, then $X \sim \text{Pois}(\lambda p)$, $Y = N - X \sim \text{Pois}(\lambda q)$, and X and Y are independent.

Proof. For a fixed $N = n$, $Y = N - X \sim \text{Bin}(n, q)$. Now,

$$P(X = x, Y = y) = \sum_{n=0}^{\infty} P(X = x, Y = y|N = n)P(N = n)$$

From this sum, only for $n = x + y$ there are non-zero probabilities, hence the sum is dropped. So,

$$\begin{aligned}
P(X=x, Y=y) &= P(X=x, Y=y|N=x+y)P(N=x+y) \\
&= \binom{x+y}{x} p^x q^y . \frac{e^{-\lambda}\lambda^{x+y}}{(x+y)!} \\
&= \frac{e^{-\lambda p}(\lambda p)^x}{x!} . \frac{e^{-\lambda q}(\lambda q)^y}{y!}.
\end{aligned}$$

And hence $X \sim \text{Pois}(\lambda p)$, $Y \sim \text{Pois}(\lambda q)$, and as the product of their PMFs form the joint PMF, X and Y are independent.

11.2 Joint Continuous Distributions

Similar to the discrete case, joint continuous distributions can be specified by their CDFs or PDFs. The following definitions are for distributions over 2 variables, but can be extended for distributions over n variables.

Defn. Joint CDF: The joint CDF $F_{X,Y}$ of two continuous random variables X and Y is a function

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y),$$

which is differentiable with respect to x and y.

Defn. Joint PDF: The derivative of the joint CDF with respect to x and y is the joint PDF function,

$$f_{X,Y}(x, y) = \frac{d^2}{dxdy} F_{X,Y}(x, y).$$

Valid joint PDFs are non-negative and integrate to 1, i.e.,

$$f_{X,Y}(x, y) \geq 0 \ \forall x, y \quad \text{and,} \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dxdy = 1.$$

The probability of events associated with arbitrary regions $A \subseteq \mathbb{R}^2$ are given by,

$$P((X, Y) \in A) = \int_A \int f_{X,Y}(x, y) dxdy.$$

Defn. Marginal PDF: For continuous random variables X, Y with joint PDF $f_{X,Y}$, the marginal PDF of X is obtained by integrating over Y,

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

Defn. Conditional PDF: For continuous random variables X, Y with joint PDF $f_{X,Y}$, the conditional PDF of X given $Y = y$ is,

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

for all y with $f_Y(y) > 0$. By convention, $f_{X|Y}(x|y) = 0$ for all y with $f_Y(y) = 0$.

We note that conditional PDFs are defined at $Y = y$ even when Y is continuous. This is interpreted as the distribution of X for values of Y over a small interval around y, i.e.,

$Y \in (y-\epsilon, y+\epsilon)$, with ϵ approaching 0 from the right.

An interesting advantage of working with joint distributions is that the joint PDF $f_{X,Y}$ can be recovered from the conditional PDF $f_{X|Y}$ and the corresponding marginal PDF f_Y, as $f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y)$. This is not possible for discrete distributions.

Prop. Bayes Rule and Law of Total Probability for joint continuous distributions hold for their PDFs:

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)}, \quad \text{for } f_X(x) > 0.$$

and,

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

The above definitions make it possible to extend these discussions easily for the hybrid case where one of X and Y are discrete whereas the other is continuous. For Bayes rule, the following definitions hold for the four possible hybrid cases.

	Discrete Y	Cont. Y
Discrete X	$P(Y=y\|X=x) = \frac{P(X=x\|Y=y)P(Y=y)}{P(X=x)}$	$f_Y(y\|X=x) = \frac{P(X=x\|Y=y)f_Y(y)}{P(X=x)}$
Cont. X	$P(Y=y\|X=x) = \frac{f_X(x\|Y=y)P(Y=y)}{f_X(x)}$	$f_{Y\|X}(y\|x) = \frac{f_{X\|Y}(x\|y)f_Y(y)}{f_X(x)}$

Similarly the definitions for the Law of Total Probability can be extended for the four hybrid cases:

	Discrete Y	Cont. Y
Discrete X	$\sum_y P(X=x\|Y=y)P(Y=y)$	$\int_{-\infty}^{\infty} P(X=x\|Y=y)f_Y(y)dy$
Cont. X	$\sum_y f_X(x\|Y=y)P(Y=y)$	$\int_{-\infty}^{\infty} f_{X\|Y}(x\|y)f_Y(y)dy$

The independence of continuous random variables can be defined in terms of the CDFs or PDFs.

Defn. The independence of continuous random variables X and Y can be defined as the factorization of the joint CDF into the product of marginal CDFs.

$$F_{X,Y}(x,y) = F_X(x)F_Y(y), \ \forall x, y.$$

Similarly the independence can be defined in terms of the PDFs,

$$f_{X,Y}(x,y) = f_X(x)f_Y(y), \ \forall x, y.$$

which is equivalent to the condition $f_{Y|X}(y|x) = f_Y(y)$ for all x, y with $f_X(x) > 0$.

Note that the marginal PDF of Y, $f_Y(y)$, is a function of y only. It is free of x. The conditional PDF $f_{Y|X}(y|x)$ can depend on x in general. Only for the case of independence is $f_{Y|X}(y|x)$ free of x.

Prop. Let the joint PDF $f_{X,Y}$ of X and Y factor as,

$$f_{X,Y}(x,y) = g(x)h(y), \ \forall x, y$$

where g and h are non-negative functions. Then X and Y are independent. Also, if either g or h are a valid PDF, then the other is a valid PDF as well, and g and h are marginal PDFs of X and Y respectively.

Proof. Let $c = \int_{-\infty}^{\infty} h(y)dy$. Then $\frac{h(y)}{c}$ is a valid PDF. Now,

$$f_{X,Y}(x, y) = cg(x).\frac{h(y)}{c},$$

from which the marginal PDF of X is,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy = cg(x)\int_{-\infty}^{\infty} \frac{h(y)}{c}dy = cg(x).$$

Similarly the marginal PDF can be derived to be $\frac{h(y)}{c}$. Thus X and Y are independent as $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

Ex. (i) Let (X, Y) be a random point in the square $\{(x, y) : x, y \in [0, 1]\}$. The joint PDF of X and Y is constant over this square and 0 outside of it. The joint PDF of this *Uniform distribution on the square* is,

$$f_{X,Y}(x, y) = \begin{cases} 1 & \text{if } x, y \in [0, 1], \\ 0 \text{ o/w.} \end{cases}$$

X and Y are marginally Unif$(0, 1)$. We can observe that X and Y here are independent.

(ii) Now let (X, Y) be a random point in the unit disk $\{(x, y) : x^2 + y^2 \leq 1\}$, with joint PDF,

$$f_{X,Y}(x, y) = \begin{cases} \frac{1}{\pi} & \text{if } x^2 + y^2 \leq 1, \\ 0 & \text{o/w.} \end{cases}$$

Now X and Y are not independent, which makes sense intuitively since larger $|X|$ constrains $|Y|$ to smaller values, and vice-versa. If we find the marginal PDF of X, we get,

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}dy = \frac{2}{\pi}\sqrt{1 - x^2}, \ -1 \leq x \leq 1.$$

And similarly the marginal PDF of Y is $f_Y(y) = \frac{2}{\pi}\sqrt{1 - y^2}$. Since the joint PDF does not factorize into the marginal PDFs, they are not independent.

Ex. Exponentials of different rates can be compared easily, as is shown in this example. Let $T_1 \sim$ Expo(λ_1) and $T_2 \sim$ Expo(λ_2) be independent. We wish to find $P(T_1 < T_2)$.

We need to integrate the joint PDF of T_1 and T_2 over the region with $t_1 > 0$, $t_2 > 0$ and $t_1 < t_2$. Hence,

$$P(T_1 < T_2) = \int_0^{\infty}\int_0^{t_2} \lambda_1 e^{-\lambda_1 t_1}\lambda_2 e^{-\lambda_2 t_2}dt_1 dt_2 = \int_0^{\infty}\left(\int_0^{t_2} \lambda_1 e^{-\lambda_1 t_1}dt_1\right)\lambda_2 e^{-\lambda_2 t_2}dt_2$$
$$= \int_0^{\infty}(1 - e^{-\lambda_1 t_2})\lambda_2 e^{-\lambda_2 t_2}dt_2 = 1 - \int_0^{\infty} \lambda_2 e^{-(\lambda_1+\lambda_2)t_2}dt_2 = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

Ex. *Cauchy PDF*: Let X and Y be i.i.d. $\mathcal{N}(0, 1)$, and let $T = X/Y$. T is arbitrary for the case of $Y = 0$ and has no effect on the distribution of T, as $P(Y = 0) = 0$.

The CDF for T can be found first.

$$F_T(t) = P(T \le t) = P\left(\frac{X}{Y} \le t\right) = P\left(\frac{X}{|Y|} \le t\right).$$

The last step is due to $\frac{X}{Y}$ and $\frac{X}{|Y|}$ being identically distributed due to the symmetry of the standard Normal distribution. As X and Y are independent,

$$\begin{aligned}
F_T(t) = P(X \le t|Y|) &= \int_{-\infty}^{\infty}\int_{-\infty}^{t|Y|} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dxdy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-y^2/2}\left(\int_{-\infty}^{t|Y|} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx\right)dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-y^2/2}\Phi(t|y|)dy = \sqrt{\frac{2}{\pi}}\int_0^{\infty} e^{-y^2/2}\Phi(ty)dy.
\end{aligned}$$

The derivative of the CDF is evaluated here. In this instance, the differentiation and integral can be interchanged.

$$\begin{aligned}
f_T(t) = F_T'(t) &= \sqrt{\frac{2}{\pi}}\int_0^{\infty}\frac{\partial}{\partial t}\left(e^{-y^2/2}\Phi(ty)\right)dy \\
&= \sqrt{\frac{2}{\pi}}\int_0^{\infty} ye^{-y^2/2}\varphi(ty)dy = \frac{1}{\pi}\int_0^{\infty} ye^{-\frac{(1+t^2)y^2}{2}}dy \\
&= \frac{1}{\pi(1+t^2)}.
\end{aligned}$$

The last step uses the substitution $u = (1+t^2)y^2/2$, $du = (1+t^2)ydy$. Thus the PDF of T is,

$$f_T(t) = \frac{1}{\pi(1+t^2)},\ t \in \mathbb{R}.$$

11.3 Expectation and Covariance

For scalar functions defined on multiple random variables, the expected value can be calculated. A commonly used summarizing measure is the covariance, which provides a way to measure the relative tendency of a random variable to increase or decrease as the other random variable increases or decreases.

Defn. Expected value of a scalar function: For a function $g : \mathbb{R}^2 \to \mathbb{R}$ defined on discrete random variables X and Y, the expected value of g is given by,

$$E(g(X,Y)) = \sum_x\sum_y g(x,y)P(X = x, Y = y).$$

If X and Y are continuous, the expected value is,

$$E(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dxdy.$$

Defn. Covariance: The covariance between random variables X and Y is defined as,

$$\text{Cov}(X,Y) = E[(X - E(X))(Y - E(Y))],$$

or equivalently,

$$\text{Cov}(X,Y) = E(XY) - E(X)E(Y).$$

A related measure that is unitless is called correlation.

Defn. Correlation:

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

For any random variables X, Y, $-1 \leq \text{Corr}(X,Y) \leq 1$.

Thm. If X and Y are independent, they are uncorrelated.

Proof. The proof is shown for continuous X and Y, and the proof for discrete random variables is similar.

$$\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x) f_Y(y) dx dy \\
&= \int_{-\infty}^{\infty} y f_Y(y) \left(\int_{-\infty}^{\infty} x f_X(x) dx\right) dy \\
&= \int_{-\infty}^{\infty} x f_X(x) dx \int_{-\infty}^{\infty} y f_Y(y) dy = E(X)E(Y) \\
\implies \text{Cov}(XY) &= 0.
\end{aligned}$$

Prop. There are several important and useful properties of covariance.

(i) $\text{Cov}(X,X) = \text{Var}(X)$, (ii) $\text{Cov}(X,Y) = \text{Cov}(Y,X)$, (iii) $\text{Cov}(X,c) = 0$, $\forall c \in \mathbb{R}$,

(iv) $\text{Cov}(aX,Y) = a\text{Cov}(C,Y)$, $\forall a \in \mathbb{R}$, (v) $\text{Cov}(X+Y,Z) = \text{Cov}(X,Z) + \text{Cov}(Y,Z)$,

(vi) $\text{Cov}(X+Y,Z+W) = \text{Cov}(X,Z) + \text{Cov}(X,W) + \text{Cov}(Y,Z) + \text{Cov}(Y,W)$,

(vii) $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$

The last property extends to n variables:

$\text{Var}(X_1 + ... + X_n) = \text{Var}(X_1) + ... + \text{Var}(X_n) + 2\sum_{i<j}\text{Cov}(X_i,X_j)$.

Avisek Gupta
Postdoctoral Fellow
IAI, TCG CREST.