Foundation Fortnight
Joint and Conditional Distributions and Expectations

Yvo Pokern

20.9.2016

Today's Program

10 – 11	Joint Distributions
11 – 11:30	Break
11:30 – 1	Transformations (Expectations start?)
1 – 2:30	Break
2:30 – 3:30	Expectations, Generating Functions
3:30 – 4	Break
4 – 5:30	Convergence, Limit Theorems

Joint Distributions

Let us first consider the bivariate case. Suppose that the two random variables X and Y share the same sample space Ω (*e.g.* the height and the weight of an individual). Then we can consider the event

$$\{\omega : X(\omega) \leq x, Y(\omega) \leq y\}$$

and define its probability, regarded as a function of the two variables x and y, to be the **joint (cumulative) distribution function** of X and Y, denoted by

$$\begin{aligned} F_{X,Y}(x,y) &= \mathrm{P}(\{\omega : X(\omega) \leq x, Y(\omega) \leq y\}) \\ &= \mathrm{P}(X \leq x, Y \leq y). \end{aligned}$$

The joint cumulative distribution function (cdf) has similar *properties* to the univariate cdf. If the function $F_{X,Y}(x, y)$ is the joint distribution function of random variables X and Y then

1. $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$ and $F_{X,Y}(\infty, \infty) = 1$ and
2. $F_{X,Y}$ is a nondecreasing function of each of its arguments x

The marginal cdfs of X and Y can be found from

$$F_X(x) = \mathrm{P}(X \leq x, Y < \infty) = F_{X,Y}(x, \infty)$$

and

$$F_Y(y) = \mathrm{P}(X < \infty, Y \leq y) = F_{X,Y}(\infty, y)$$

respectively.

The joint cumulative distribution function (cdf) has similar *properties* to the univariate cdf. If the function $F_{X,Y}(x,y)$ is the joint distribution function of random variables X and Y then

1. $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$ and $F_{X,Y}(\infty, \infty) = 1$ and
2. $F_{X,Y}$ is a nondecreasing function of each of its arguments x

The marginal cdfs of X and Y can be found from

$$F_X(x) = \mathrm{P}(X \leq x, Y < \infty) = F_{X,Y}(x, \infty)$$

and

$$F_Y(y) = \mathrm{P}(X < \infty, Y \leq y) = F_{X,Y}(\infty, y)$$

respectively.

We already know in the univariate case that $P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$. Similarly, we find in the bivariate case that

$$P(x_1 < X \leq x_2, \;\; y_1 < Y \leq y_2) =$$
$$F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1).$$

Example (0)

Consider the function

$$F_{X,Y}(x, y) = x^2y + y^2x - x^2y^2\,, \quad 0 \leq x \leq 1, 0 \leq y \leq 1\,.$$

Show that $F_{X,Y}$ is a joint cdf of two continuous random variables, X and Y. Find the marginal cdfs of X and Y. Also find $P(0 \leq X \leq \frac{1}{2}, 0 \leq Y \leq \frac{1}{2})$.

We already know in the univariate case that $P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$. Similarly, we find in the bivariate case that

$$P(x_1 < X \leq x_2, \;\; y_1 < Y \leq y_2) =$$

$$F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1).$$

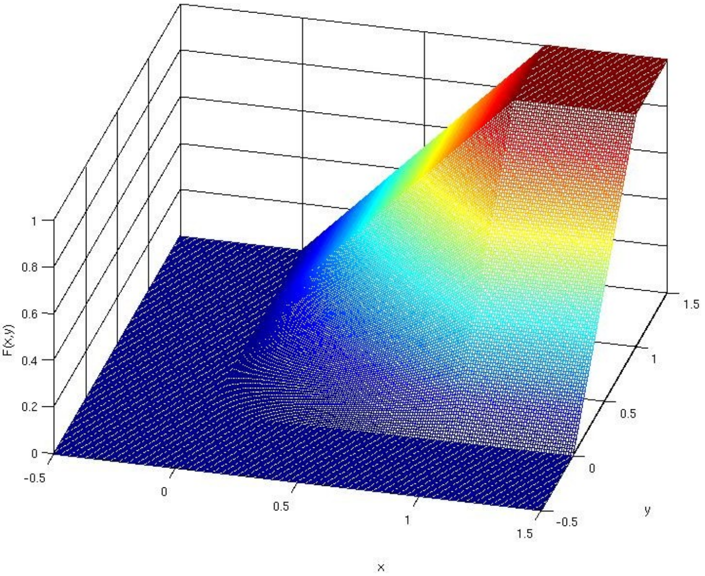
Example (1)

Consider the function

$$F_{X,Y}(x, y) = x^2 y + y^2 x - x^2 y^2 \, , \quad 0 \leq x \leq 1, 0 \leq y \leq 1 \, .$$

Show that $F_{X,Y}$ is a joint cdf of two continuous random variables, X and Y. Find the marginal cdfs of X and Y. Also find $P(0 \leq X \leq \frac{1}{2}, 0 \leq Y \leq \frac{1}{2})$.

$$F_{X,Y}(x,y) = x^2y + y^2x - x^2y^2\,, \quad 0 \le x \le 1,\, 0 \le y \le 1\,.$$



Joint and Marginal PMF I

In many cases of interest, X and Y take only values in a discrete set. Then $F_{X,Y}$ is a step function in each variable separately and we consider the **joint probability mass function**

$$p_{X,Y}(x_i, y_j) = \mathrm{P}(X = x_i, Y = y_j).$$

It is often convenient to represent a discrete **bivariate** distribution – a joint distribution of two variables – by a *2-way table.* In general, the entries in the table are the *joint probabilities* $p_{X,Y}(x, y)$, while the row and column totals give the *marginal* probabilities $p_X(x)$ and $p_Y(y)$. As always, the total probability is 1.

Joint and Marginal PMF II

Example (2)

Consider three independent tosses of a fair coin. Let $X =$ 'number of heads in first and second toss' and $Y =$ 'number of heads in second and third toss'. Give the probabilities for any combination of possible outcomes of X and Y in a two-way table and obtain the marginal pmfs of X and Y.

Joint and Marginal PMF III

In general, from the joint distribution we can use the law of total probability to obtain the **marginal pmf** of Y as

$$p_Y(y_j) = \mathrm{P}(Y = y_j) = \sum_{x_i} \mathrm{P}(X = x_i,\ Y = y_j) = \sum_{x_i} p_{X,Y}(x_i, y_j)\,.$$

Similarly, the **marginal pmf** of X is given by

$$p_X(x_i) = \sum_{y_j} p_{X,Y}(x_i, y_j)\,.$$

The marginal distribution is thus the distribution of just one of the variables.
The joint cdf can be written as

$$F_{X,Y}(x, y) = \sum_{x_i \le x} \sum_{y_j \le y} p_{X,Y}(x_i, y_j)\,.$$

Independence I

The random variables X and Y, defined on the sample space Ω with probability function P, are **independent** if the events

$$\{X = x_i\} \text{ and } \{Y = y_j\}$$

are *independent events*, for all possible values x_i and y_j. Thus X and Y are independent if,

$$p_{X,Y}(x_i, y_j) = \mathrm{P}(X = x_i, Y = y_j) = p_X(x_i)p_Y(y_j)$$

for all x_i, y_j.

This implies that $\mathrm{P}(X \in A, Y \in B) = \mathrm{P}(X \in A)\mathrm{P}(Y \in B)$ for all sets A and B, so that the two events
$\{\omega : X(\omega) \in A\}, \{\omega : Y(\omega) \in B\}$ are independent.
(*Exercise:* prove this.)

Independence II

NB: If x is such that $p_x(x) = 0$, then $p_{x,Y}(x, y_j) = 0$ for all y_j and the above factorisation holds automatically. Thus it does not matter whether we require the factorisation for all *possible* x_i, y_j *i.e.* those with positive probability, or all *real* x, y. (That is, $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ for all x, y would be an equivalent definition of independence.)

If X, Y are independent then the entries in the two way table are the products of the marginal probabilities. In Problem 3.8.1 in Rice we see that X and Y are *not* independent.

Motivation

These are defined for random variables by analogy with conditional probabilities of events. Consider the conditional probability

$$\begin{aligned} \mathrm{P}(X = x_i \mid Y = y_j) &= \frac{\mathrm{P}(X = x_i,\ Y = y_j)}{\mathrm{P}(Y = y_j)} \\ &= \frac{p_{X,Y}(x_i, y_j)}{p_Y(y_j)} \end{aligned}$$

as a function of x_i, for fixed y_j. Then this is a probability mass function – it is non-negative and

$$\sum_{x_i} \frac{p_{X,Y}(x_i, y_j)}{p_Y(y_j)} = \frac{1}{p_Y(y_j)} \underbrace{\sum_{x_i} p_{X,Y}(x_i, y_j)}_{p_Y(y_j)} = 1,$$

and it gives the probabilities for observing $X = x_i$ given that we already know $Y = y_j$.

Definition

We therefore *define* the **conditional probability distribution** of X given $Y = y_j$ as

$$\boxed{p_{X|Y}(x_i|y_j) = \frac{p_{X,Y}(x_i, y_j)}{p_Y(y_j)}}$$

Conditioning on $Y = y_j$ can be compared to selecting a subset of the population, i.e. only those individuals where $Y = y_j$. The conditional distribution $p_{X|Y}$ of X given $Y = y_j$ then describes the distribution of X within this subgroup.

Multiplication Rule

$$\boxed{p_{X,Y}(x_i, y_j) = p_{X|Y}(x_i|y_j)p_Y(y_j)}$$

which can be used to find a bivariate pmf when we known one marginal distribution and one conditional distribution. Note that

if X and Y are *independent* then $p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_i)$ so that $p_{X|Y}(x_i|y_j) = p_X(x_i)$ *i.e.* the *conditional* distribution is the same as the *marginal* distribution.

Indpendence, Marginal and Conditional

In general, X and Y are independent *if and only if* the conditional distribution of X given $Y = y_j$ is the same as the marginal distribution of X *for all* y_j. (This condition is equivalent to $p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_j)$ for all x_i, y_i, above).
The conditional distribution of Y given $X = x_i$ is defined similarly.

Question 3.8.1

Obtain the conditional pmf of X given $Y = y$. Use this conditional distribution to verify that X and Y are not independent. □

Univariate Setup

In this section we will see how to derive the distribution of transformed random variables. This is useful because many statistics applied to data analysis (*e.g.* test statistics) are transformations of the sample variables.

Suppose that we have a sample space Ω, a probability function P on Ω, a random variable $X : \Omega \to \mathbb{R}$, and a function $\phi : \mathbb{R} \to \mathbb{R}$.
Recall $Y = \phi(X) : \Omega \to \mathbb{R}$ is defined by $Y(\omega) = \phi(X)(\omega) = \phi(X(\omega))$. Since $Y = \phi(X)$ is a random variable it also has a probability distribution, which can be determined either directly from P or via the distribution of X.

Discrete Case

$$
\begin{aligned}
\mathrm{P}(Y = y) = \mathrm{P}(\{\omega : \phi(X(\omega)) = y\}) &= \sum_{\{\omega:\phi(X(\omega))=y\}} \mathrm{P}(\{\omega\}) \\
&= \sum_{\{x:\phi(x)=y\}} \mathrm{P}(\{\omega : X(\omega) = x\}) \\
&= \sum_{\{x:\phi(x)=y\}} p_X(x).
\end{aligned}
$$

So, for example

$$
\begin{aligned}
\mathrm{E}[Y] &= \sum_{\omega} \phi(X(\omega))\mathrm{P}(\{\omega\}) && \text{with respect to P on } \Omega \\
&= \sum_{x} \phi(x) p_X(x) && \text{with respect to distribution of } X \\
&= \sum_{y} y p_{\phi(X)}(y) && \text{with respect to distribution of } \phi(X)
\end{aligned}
$$

Discrete Case

$$\begin{aligned}
\mathrm{P}(Y = y) = \mathrm{P}(\{\omega : \phi(X(\omega)) = y\}) &= \sum_{\{\omega:\phi(X(\omega))=y\}} \mathrm{P}(\{\omega\}) \\
&= \sum_{\{x:\phi(x)=y\}} \mathrm{P}(\{\omega : X(\omega) = x\}) \\
&= \sum_{\{x:\phi(x)=y\}} p_X(x).
\end{aligned}$$

So, for example

$$\begin{aligned}
\mathrm{E}[Y] &= \sum_{\omega} \phi(X(\omega))\mathrm{P}(\{\omega\}) \quad \text{with respect to P on } \Omega \\
&= \sum_{x} \phi(x) p_X(x) \quad \text{with respect to distribution of } X \\
&= \sum_{y} y p_{\phi(X)}(y) \quad \text{with respect to distribution of } \phi(X)
\end{aligned}$$

Continuous & Discrete Example

Example (3)

Consider two independent throws of a fair die. Let X be the sum of the numbers that show up. Give the distribution of X. Now consider the transformation $Y = (X - 7)^2$. Derive the distribution of Y.

Continuous case: Similar to above, we may sometimes work out the distribution of a transformation directly even in the continuous case as illustrated in the next example.

Example (4)

Let X be a continuous random variable with the density

$$f_X(x) = \frac{\lambda}{2} \exp(-\lambda|x - \mu|) \quad x \in \mathbb{R}, \lambda > 0.$$

(This is the two–sided exponential distribution centred at μ.) Find the distribution of $Y = |X - \mu|$.

The probability integral transform

The Probability Integral Transform

Let X be a continuous random variable with distribution function $F(x)$, strictly increasing over the range of X. Let $Y = F(X)$ (*i.e.* Y is the distribution function of X evaluated at the random variable X). Then Y is uniformly distributed on $(0, 1)$.

This result has an important application to the simulation of random variables: to generate an observation from the continuous distribution F, first generate a uniform pseudo-random number, Y, and then compute $X = F^{-1}(Y)$. Then X will be an observation from F.

Example (5)

Use the probability integral transform to generate values from an $\text{Exp}(\lambda)$ distribution, given a uniform random number generator.

Proof of Result

First note that $0 \le Y \le 1$ (since F is a distribution function). Let $0 \le y \le 1$; then

$$\mathrm{P}(Y \le y) = \mathrm{P}(F(X) \le y) = \mathrm{P}(X \le F^{-1}(y)) = F(F^{-1}(y)) = y\,.$$

The third step follows because F is a strictly increasing function.

By differentiation, it follows that Y has pdf $f(y) = 1$ on $(0, 1)$, which is the density of the uniform distribution on $(0, 1)$. $\square$

General Case: ϕ Increasing

In general, suppose that $Y = \phi(X)$ where ϕ is a strictly increasing and differentiable function. Then, following the same method as above,

$$F_Y(y) = \mathrm{P}(\phi(X) \leq y) = \mathrm{P}(X \leq \phi^{-1}(y)) = F_X(\phi^{-1}(y)).$$

Then, differentiating, Y has density

$$f_Y(y) = f_X(\phi^{-1}(y))\,\frac{d}{dy}\phi^{-1}(y) = f_X(x)\frac{dx}{dy}\bigg|_{x=\phi^{-1}(y)},$$

where the index $x = \phi^{-1}(y)$ means that any x in the formula has to be replaced by the inverse $\phi^{-1}(y)$ because $f_Y(y)$ is a function of y.

General Case: ϕ Increasing

In general, suppose that $Y = \phi(X)$ where ϕ is a strictly increasing and differentiable function. Then, following the same method as above,

$$F_Y(y) = \mathrm{P}(\phi(X) \leq y) = \mathrm{P}(X \leq \phi^{-1}(y)) = F_X(\phi^{-1}(y)).$$

Then, differentiating, Y has density

$$f_Y(y) = f_X(\phi^{-1}(y))\,\frac{d}{dy}\phi^{-1}(y) = f_X(x)\frac{dx}{dy}\bigg|_{x=\phi^{-1}(y)},$$

where the index $x = \phi^{-1}(y)$ means that any x in the formula has to be replaced by the inverse $\phi^{-1}(y)$ because $f_Y(y)$ is a function of y.

General Case: ϕ decreasing

Similarly, if ϕ is decreasing then

$$F_Y(y) = \mathrm{P}(\phi(X) \le y) = \mathrm{P}(X \ge \phi^{-1}(y)) = 1 - F_X(\phi^{-1}(y))$$

so that

$$f_Y(y) = -f_X(x)\frac{dx}{dy}\bigg|_{x=\phi^{-1}(y)}.$$

In the first case $dy/dx = d\phi(x)/dx$ is positive (since ϕ is increasing), in the second it is negative (since ϕ is decreasing) so either way the transformation formula is

$$\boxed{f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|_{x=\phi^{-1}(y)}}$$

General Case: ϕ decreasing

Similarly, if ϕ is decreasing then

$$F_Y(y) = \mathrm{P}(\phi(X) \le y) = \mathrm{P}(X \ge \phi^{-1}(y)) = 1 - F_X(\phi^{-1}(y))$$

so that

$$f_Y(y) = -f_X(x) \frac{dx}{dy}\bigg|_{x=\phi^{-1}(y)}.$$

In the first case $dy/dx = d\phi(x)/dx$ is positive (since ϕ is increasing), in the second it is negative (since ϕ is decreasing) so either way the transformation formula is

$$\boxed{f_Y(y) = f_X(x) \left|\frac{dx}{dy}\right|_{x=\phi^{-1}(y)}}$$

The Transformed density is a pdf

Recall that $\int_{-\infty}^{\infty} f_X(x)dx = 1$. Changing variable to $y = \phi(x)$ we have, for ϕ increasing,

$$1 = \int_{-\infty}^{\infty} \left\{ f_X(x) \frac{dx}{dy} \right\}_{x=\phi^{-1}(y)} dy$$

so that $f_X(x) \left| \frac{dx}{dy} \right|$ is a valid pdf. Similarly for ϕ decreasing.

Transformation equivalent to change of coordinates

The rule for change of variable $y = \phi(x)$ in an integral is

$$\int_a^b g(x) f_X(x)\,dx = \int_{\phi(a)}^{\phi(b)} g(\phi^{-1}(y))\,\underbrace{f_X(\phi^{-1}(y))\,\frac{d\phi^{-1}}{dy}(y)}_{|\cdot|=f_Y(y)}\,dy.$$

Since this holds for any (measurable, $L^1_{f_X}$-integrable) function g, it reduces to the transformation formula upon identifying $\left|f_X(\phi^{-1}(y))\,\frac{d\phi^{-1}}{dy}(y)\right|$ as $f_Y(y)$.
Note: this works for both ϕ increasing and ϕ decreasing - the negative sign of $\frac{d\phi^{-1}}{dy}(y)$ is compensated by the change of sign induced by swapping integration boundaries.

Example (6)

Consider $X \sim$Uniform$[-\frac{\pi}{2}, \frac{\pi}{2}]$, i.e.

$$f_X(x) = \begin{cases} \frac{1}{\pi} & -\frac{\pi}{2} \leq x \leq \frac{\pi}{2} \\ 0 & \text{otherwise.} \end{cases}$$

Derive the density of $Y = \tan(X)$.

General Case: ϕ neither increasing nor decreasing

When ϕ is a *many-to-one* function we use the generalised formula $f_Y(y) = \sum f_X(x) \left|\frac{dx}{dy}\right|$, where the summation is over the set $\{x : \phi(x) = y\}$. That is, we add up the contributions to the density at y from all x values which map to y.

Example (7)

Suppose that $f_X(x) = 2x$ on $(0, 1)$ and let $Y = (X - \frac{1}{2})^2$. Obtain the pdf of Y.

Bivariate Case I

For the bivariate case we consider two random variables X, Y with joint density $f_{X,Y}(x, y)$. What is the joint density of transformations $U = u(X, Y)$, $V = v(X, Y)$ where $u(\cdot, \cdot)$ and $v(\cdot, \cdot)$ are functions from $\mathbb{R}^2$ to $\mathbb{R}$, such as the ratio X/Y or the sum $X + Y$?
In order to use the following generalisation of the previous method, we need to assume that u, v are such that each pair (x, y) defines a unique (u, v) and conversely, so that $u = u(x, y)$ and $v = v(x, y)$ are differentiable and invertible.
The formula that gives the joint density of U, V is similar to the univariate case but the derivative, as we used it above, now has to be replaced by the *Jacobian* $J(u, v)$ of this transformation.

Bivariate Case II

The result is that $U = u(X, Y), V = v(X, Y)$ have joint density

$$\boxed{f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v))|J(x, y)|_{\substack{x=x(u,v)\\ y=y(u,v)}}}$$

Again, the index $\substack{x=x(u,v)\\ y=y(u,v)}$ means that the x, y have to be replaced by the suitable transformations involving u, v only. But how do we get the Jacobian $J(x, y)$? It is actually the determinant of the *matrix of partial derivatives* :

$$J(x, y) = \left|\frac{\partial(x, y)}{\partial(u, v)}\right| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

We finally take its absolute value, $|J(x, y)|$.

Computing Bivariate Transformations

1. Obtain the inverse transformation $x = x(u, v), y = y(u, v)$, compute the matrix of partial derivatives $\partial(x, y)/\partial(u, v)$ and then its determinant and absolute value.
2. Alternatively find the determinant $J(u, v)$ from the matrix of partial derivatives of (u, v) with respect to (x, y) and then its absolute value and invert this.

The two methods are equivalent since

$$\frac{\partial(x, y)}{\partial(u, v)} = \left\{\frac{\partial(u, v)}{\partial(x, y)}\right\}^{-1}$$

Which way to choose in a specific case will depend on which functions are easier to derive. But note that the inverse transformations $x = x(u, v)$ and $y = y(u, v)$ are required anyway so that the first approach is often preferable.

Computing Bivariate Transformations

1. Obtain the inverse transformation $x = x(u, v), y = y(u, v)$, compute the matrix of partial derivatives $\partial(x, y)/\partial(u, v)$ and then its determinant and absolute value.
2. Alternatively find the determinant $J(u, v)$ from the matrix of partial derivatives of (u, v) with respect to (x, y) and then its absolute value and invert this.

The two methods are equivalent since

$$\frac{\partial(x, y)}{\partial(u, v)} = \left\{ \frac{\partial(u, v)}{\partial(x, y)} \right\}^{-1}$$

Which way to choose in a specific case will depend on which functions are easier to derive. But note that the inverse transformations $x = x(u, v)$ and $y = y(u, v)$ are required anyway so that the first approach is often preferable.

Bivariate Case: Examples

Example (8)

Let X and Y be two independent exponential variables with $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\mu)$. Find the distribution of $U = X/Y$

Example (9)

Consider two independent and identically distributed random variables X and Y having a uniform distribution on $[0, 2]$. Derive the joint density of $Z = X/Y$ and $W = Y$, stating the area where this density is positive. Are Z and W independent? Obtain the marginal density of $Z = X/Y$.

Sums of Random Variables

The distribution of a sum $Z = X + Y$ of two (not necessarily independent) random variables X and Y can be derived directly as follows.
In the discrete case note that the marginal distribution of Z is

$$\mathrm{P}(Z = z) = \sum_x \mathrm{P}(X = x, Z = z) = \sum_x \mathrm{P}(X = x, Y = z - x)$$

That is,

$$\boxed{p_Z(z) = \sum_x p_{X,Y}(x, z - x)}$$

Analogously, in the continuous case we get

$$\boxed{f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\, dx}$$

Sums of Random Variables

The distribution of a sum $Z = X + Y$ of two (not necessarily independent) random variables X and Y can be derived directly as follows.
In the discrete case note that the marginal distribution of Z is

$$\mathrm{P}(Z = z) = \sum_x \mathrm{P}(X = x, Z = z) = \sum_x \mathrm{P}(X = x, Y = z - x)$$

That is,

$$\boxed{p_Z(z) = \sum_x p_{X,Y}(x, z - x)}$$

Analogously, in the continuous case we get

$$\boxed{f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\, dx}$$

Sums of Random Variables: Examples

Example (10)

Let X and Y be two positive random variables with joint pdf

$$f_{X,Y}(x,y) = xye^{-(x+y)}\,, \;\; x, y > 0\,.$$

Derive and name the distribution of their sum $Z = X + Y$.

IF X and Y are independent, one can do something much more efficient exloiting moment generating functions. Stay tuned!

Transformations for more than 2 variables

These ideas extend in a straightforward way to the case of more than two variables. The general problem is to find the distribution of $\boldsymbol{Y} = \phi(\boldsymbol{X})$, where $\boldsymbol{Y}$ is $s \times 1$ and $\boldsymbol{X}$ is $r \times 1$, from the known distribution of $\boldsymbol{X}$. Here $\boldsymbol{X}$ is the random *vector*

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ X_r \end{pmatrix} .$$

Transformation of Vector-Valued Random Variables

Discrete case. $p_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum p_{\boldsymbol{X}}(\boldsymbol{x})$, where the summation is over the set $\{\boldsymbol{x} : \phi(\boldsymbol{x}) = \boldsymbol{y}\}$. That is, we just add up the probabilities of all $\boldsymbol{x}$-values that give $\phi(\boldsymbol{x}) = \boldsymbol{y}$.

Continuous case. Case (i): ϕ is a one-to-one transformation (so that $s = r$). Then the rule is

$$\boxed{f_Y(y) = f_X(x(y))\,|J(x)|_{x=x(y)}}$$

where $J(x) = \left|\frac{dx}{dy}\right|$ is the Jacobian of transformation. Here $\frac{dx}{dy}$ is the matrix of partial derivatives $\left(\frac{dx}{dy}\right)_{ij} = \frac{\partial x_i}{\partial y_j}$.

Case (ii): $s < r$. First transform the s-vector Y to the r-vector Y', where $Y'_i = Y_i$, $i = 1, \ldots, s$, and the other $r - s$ random variables Y'_i, $i = s+1, \ldots, r$, are chosen for convenience. Now find the density of Y' as in case (i) and then integrate out $Y'_{s+1}, \ldots, Y'_r$ to obtain the marginal density of Y, as required.

Transformation of Vector-Valued Random Variables

Discrete case. $p_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum p_{\boldsymbol{X}}(\boldsymbol{x})$, where the summation is over the set $\{\boldsymbol{x} : \phi(\boldsymbol{x}) = \boldsymbol{y}\}$. That is, we just add up the probabilities of all $\boldsymbol{x}$-values that give $\phi(\boldsymbol{x}) = \boldsymbol{y}$.

Continuous case. Case (i): ϕ is a one-to-one transformation (so that $s = r$). Then the rule is

$$\boxed{f_Y(y) = f_X(x(y)) \left|J(x)\right|_{x=x(y)}}$$

where $J(x) = \left|\frac{dx}{dy}\right|$ is the Jacobian of transformation. Here $\frac{dx}{dy}$ is the matrix of partial derivatives $\left(\frac{dx}{dy}\right)_{ij} = \frac{\partial x_i}{\partial y_j}$.

Case (ii): $s < r$. First transform the s-vector Y to the r-vector Y', where $Y'_i = Y_i$, $i = 1, \ldots, s$, and the other $r - s$ random variables Y'_i, $i = s + 1, \ldots, r$, are chosen for convenience. Now find the density of Y' as in case (i) and then integrate out $Y'_{s+1}, \ldots, Y'_r$ to obtain the marginal density of Y, as required.

Transforming vector-valued RVs: non-monotonic case

Case (iii): $s = r$ but $\phi(\cdot)$ is not monotonic. Then there will generally be more than one value of $\boldsymbol{x}$ corresponding to a given $\boldsymbol{y}$ and we need to add the probability contributions from all relevant $\boldsymbol{x}$s.

Example (11)

Suppose that $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$, where $\boldsymbol{A}$ is an $r \times r$ nonsingular matrix. Then $f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{A}^{-1}\boldsymbol{y})|\boldsymbol{A}|_{+}^{-1}$, where $|\boldsymbol{A}|_{+}$ denotes the absolute value of the determinant of $\boldsymbol{A}$.

Order Statistics

Order statistics are a special kind of transformation of the sample variables. Their joint and marginal distributions can be derived by combinatorial considerations.

Suppose that $X_1, \ldots, X_n$ are independent with common density f_X. Denote the ordered values by $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$. What is the distribution F_r of $X_{(r)}$?

In particular, $X_{(n)} = \max\,(X_1, \ldots, X_n)$ is the **sample maximum** and $X_{(1)} = \min\,(X_1, \ldots, X_n)$ is the **sample minimum**.

Distribution of Maximum and Minimum

To find the distribution of $X_{(n)}$, note that $\{X_{(n)} \leq x\}$ and $\{\text{all } X_i \leq x\}$ are the same event – and so have the same probability!

$$\begin{aligned} F_n(x) = \mathrm{P}(X_{(n)} \leq x) &= \mathrm{P}(\text{all } X_i \leq x) \\ &= \mathrm{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) \\ &= \{F_X(x)\}^n \end{aligned}$$

since the X_i are independent with the same distribution function F_X. Thus

$$\boxed{F_n(x) = \{F_X(x)\}^n}$$

Furthermore, differentiating this expression we see that the density f_n of $X_{(n)}$ is

$$\boxed{f_n(x) = n\{F_X(x)\}^{n-1} f_X(x)}$$

Distribution of Maximum and Minimum

To find the distribution of $X_{(n)}$, note that $\{X_{(n)} \leq x\}$ and $\{\text{all } X_i \leq x\}$ are the same event – and so have the same probability!

$$\begin{aligned} F_n(x) = \mathrm{P}(X_{(n)} \leq x) &= \mathrm{P}(\text{all } X_i \leq x) \\ &= \mathrm{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) \\ &= \{F_X(x)\}^n \end{aligned}$$

since the X_i are independent with the same distribution function F_X. Thus

$$\boxed{F_n(x) = \{F_X(x)\}^n}$$

Furthermore, differentiating this expression we see that the density f_n of $X_{(n)}$ is

$$\boxed{f_n(x) = n\{F_X(x)\}^{n-1} f_X(x)}$$

Distribution of $X_{(r)}$

For dx sufficiently small we have

$$\mathrm{P}(x < X_{(r)} \le x + dx) = \mathrm{P}\left(\begin{array}{l} r-1 \text{ values } X_i \text{ such that } X_i \le x, \text{ and} \\ \text{one value in } (x, x+dx]\,, \text{ and} \\ n-r \text{ values such that } X_i > x+dx \end{array}\right)$$

$$\simeq \frac{n!}{(r-1)!(n-r)!}\{F_X(x)\}^{r-1} f_X(x)dx\,\{1 - F_X(x+dx)\}^{n-r}$$

Recalling that $f_r(x) = \lim_{dx\to 0} \mathrm{P}(x < X_{(r)} \le x + dx)/dx$, dividing both sides of the above expression by dx and letting $dx \to 0$ we obtain the density function of the rth order statistic $X_{(r)}$ as

$$\boxed{f_r(x) = \frac{n!}{(r-1)!(n-r)!}\{F_X(x)\}^{r-1}\{1 - F_X(x)\}^{n-r} f_X(x)}$$

In particular, this formula gives the previous densities when $r = n$ and $r = 1$.

Distribution of $X_{(r)}$

For dx sufficiently small we have

$$\mathrm{P}(x < X_{(r)} \leq x + dx) = \mathrm{P}\left(\begin{array}{l} r-1 \text{ values } X_i \text{ such that } X_i \leq x, \text{ and} \\ \text{ one value in } (x, x+dx]\,, \text{ and} \\ n-r \text{ values such that } \ X_i > x+dx \end{array} \right)$$

$$\simeq \frac{n!}{(r-1)!(n-r)!} \{F_X(x)\}^{r-1} f_X(x) dx \,\{1 - F_X(x+dx)\}^{n-r}$$

Recalling that $f_r(x) = \lim_{dx \to 0} \mathrm{P}(x < X_{(r)} \leq x + dx)/dx$, dividing both sides of the above expression by dx and letting $dx \to 0$ we obtain the density function of the rth order statistic $X_{(r)}$ as

$$\boxed{f_r(x) = \frac{n!}{(r-1)!(n-r)!} \{F_X(x)\}^{r-1} \{1 - F_X(x)\}^{n-r} f_X(x)}$$

In particular, this formula gives the previous densities when $r = n$ and $r = 1$.

Joints density of $X_{(1)}$ and $X_{(2)}$

$$f_{1,n}(x,y) = n(n-1)\{F_X(y) - F_X(x)\}^{n-2} f_X(x) f_X(y)$$

since there are $n(n-1)$ ways of choosing the variables to be the largest and smallest, and $n-2$ variables must lie in (x, y).

Applications

Order statistics are widely used in nonparametric statistics (that is, statistical methods in which we make no assumptions about the form of the population distribution). They are also used extensively in reliability theory and quality assurance. Consider for instance a system that consists of four components with independent lifetimes. How do we model the lifetime of the whole system if

1. the system fails whenever one component fails?
2. the system only fails if all components fail?

Afternoon

- Expectations in more complicated contexts (Covariance, prediction, moment generating functions)
- Law of Large Numbers**!!**
- Central Limit Theorem **!!!**