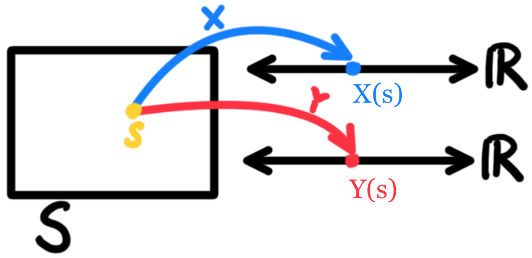
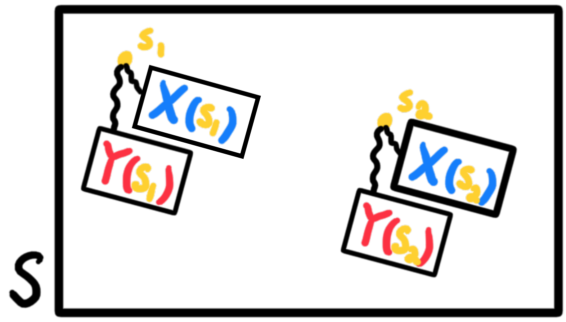
Chapter 7. Jointly Distributed Random Variables

Suppose we have two random variables X and Y defined on a sample space S with probability measure $\mathrm{P}(E)$ (where $E \subseteq S$ is an event).



or



Note that S could be the set of outcomes from two 'experiments', and the sample space points be two-dimensional; then perhaps X could relate to the first experiment, and Y to the second. But this is not necessarily the case and X and Y can be more intertwined.

From before we know to define the *marginal* probability distributions P_X and P_Y by

$$\mathrm{P}_X(B) = \mathrm{P}(X^{-1}(B)), \quad B \subseteq \mathbb{R}.$$

Reminder, $X^{-1}(B) = \{s \in S : X(s) \in B\}$

We now define the **joint probability distribution**:

Definition 7.0.1. *Given a pair of random variables, X and Y, we define the* **joint probability distribution** P_{XY} *as follows:*

$$\begin{aligned} \mathrm{P}_{XY}(B_X, B_Y) &= \mathrm{P}\left(X^{-1}(B_X) \cap Y^{-1}(B_Y)\right), \\ &= \mathrm{P}\left(\{s \in S : X(s) \in B_X \text{ and } Y(s) \in B_Y\}\right), \qquad B_X, B_Y \subseteq \mathbb{R}. \end{aligned}$$

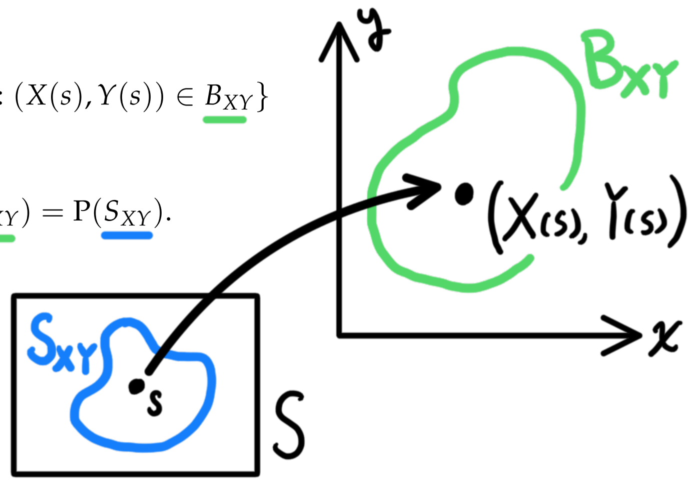
So $\mathrm{P}_{XY}(B_X, B_Y)$, the probability that $X \in B_X$ **and** $Y \in B_Y$, is given by the probability of the set of outcomes in the sample space that simultaneously get mapped into B_X by X **and** into B_Y by Y.

More generally, for some two-dimensional region $B_{XY} \subseteq \mathbb{R}^2$, find the collection of sample space elements (i.e. the event)

$$S_{XY} = \{s \in S : (X(s), Y(s)) \in B_{XY}\}$$

and define

$$\mathrm{P}_{XY}(B_{XY}) = \mathrm{P}(S_{XY}).$$



7.0.1 Joint Cumulative Distribution Function

We define the joint cumulative distribution as follows:

Definition 7.0.2. *Given a pair of random variables, X and Y, the joint cumulative distribution function is defined as*

$$F_{XY}(x,y) = \mathrm{P}_{XY}(X \le x, Y \le y), \quad x, y \in \mathbb{R}.$$

It is easy to check that the marginal cdfs for X and Y can be recovered by

$$\begin{aligned} F_X(x) &= F_{XY}(x,\infty), \quad x \in \mathbb{R}, \\ F_Y(y) &= F_{XY}(\infty,y), \quad y \in \mathbb{R}. \end{aligned}$$

7.0.2 Properties of Joint CDF F_{XY}

For F_{XY} to be a valid cdf, we need to make sure the following conditions hold.

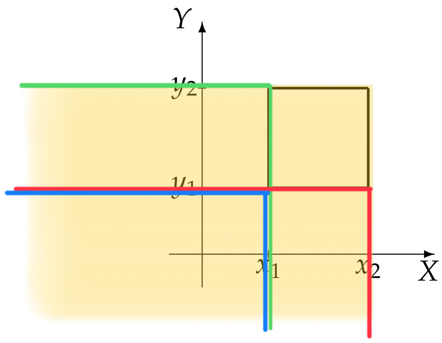
1. $0 \le F_{XY}(x,y) \le 1, \forall x, y \in \mathbb{R}$;
2. Monotonicity:

   For any fixed $y \in \mathbb{R}$, $\quad x_1 < x_2 \Rightarrow F_{XY}(x_1,y) \le F_{XY}(x_2,y)$
   For any fixed $x \in \mathbb{R}$, $\quad y_1 < y_2 \Rightarrow F_{XY}(x,y_1) \le F_{XY}(x,y_2)$

3. $\forall x, y \in \mathbb{R}$,

   $F_{XY}(x,-\infty) = 0, F_{XY}(-\infty,y) = 0$ and $F_{XY}(\infty,\infty) = 1$.

Suppose we are interested in whether the random variable pair (X,Y) lie in the interval Cartesian product $(x_1,x_2] \times (y_1,y_2]$; that is, if $x_1 < X \le x_2$ *and* $y_1 < Y \le y_2$.



First note that $\mathrm{P}_{XY}(x_1 < X \le x_2,\ Y \le y) = F_{XY}(x_2,y) - F_{XY}(x_1,y)$.

It is then easy to see that $\mathrm{P}_{XY}(x_1 < X \le x_2,\ y_1 < Y \le y_2)$ is given by

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

7.0.3 Joint Probability Mass Functions

Definition 7.0.3. *If X and Y are both discrete random variables, then we can define the* **joint probability mass function** *as*

$$p_{XY}(x,y) = \mathrm{P}_{XY}(X = x, Y = y), \quad x, y \in \mathbb{R}.$$

We can recover the marginal pmfs p_X and p_Y since, by the law of total probability, $\forall x, y \in \mathbb{R}$,

$$p_X(x) = \sum_{y \in \mathbb{Y}} p_{XY}(x,y), \quad p_Y(y) = \sum_{x \in \mathbb{X}} p_{XY}(x,y).$$

Properties of Joint PMFs

For p_{XY} to be a valid pmf, we need to make sure the following conditions hold.

1. $0 \le p_{XY}(x,y) \le 1, \forall x, y \in \mathbb{R}$;
2. $\sum_{y \in \mathbb{Y}} \sum_{x \in \mathbb{X}} p_{XY}(x,y) = 1$.

7.0.4 Joint Probability Density Functions

On the other hand, if $\exists f_{XY} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that for any $B_{XY} \subseteq \mathbb{R} \times \mathbb{R}$,

$$\mathrm{P}_{XY}(B_{XY}) = \int_{(x,y) \in B_{XY}} f_{XY}(x,y) dx dy,$$

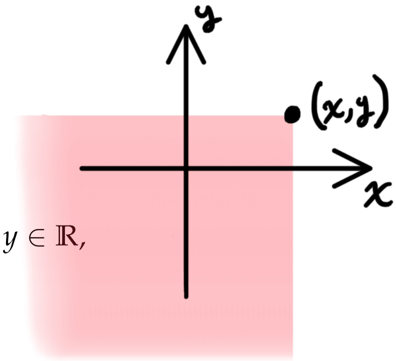
then we say X and Y are **jointly continuous** and we refer to f_{XY} as the **joint probability density function** of X and Y.

Exactly as with a single continuous random variable, we have the following helpful interpretation of the joint pdf:

$f_{XY}(x,y)\, dx\, dy$ is the probability that X is between x and $x + dx$
and Y is between y and $y + dy$

(where we understand this statement in the limit as dx and dy go to zero).

For continuous random variables, we have



$$F_{XY}(x,y) = \int_{t=-\infty}^{y} \int_{s=-\infty}^{x} f_{XY}(s,t)\, ds\, dt, \quad x, y \in \mathbb{R},$$

Definition 7.0.4. *By the Fundamental Theorem of Calculus we can identify the joint pdf as*

$$f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y).$$

Furthermore, we can recover the marginal densities f_X and f_Y from the joint CDF, e.g.,

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) = \frac{d}{dx} F_{XY}(x, \infty) \\ &= \frac{d}{dx} \int_{y=-\infty}^{\infty} \int_{s=-\infty}^{x} f_{XY}(s,y)\, ds\, dy, \end{aligned}$$

using the Fundamental Theorem of Calculus. Through a symmetric argument for Y, we get

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

Properties of Joint PDFs

For f_{XY} to be a valid pdf, we need to make sure the following conditions hold.

1. $f_{XY}(x,y) \geq 0, \forall x, y \in \mathbb{R}$;
2. $\int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} f_{XY}(x,y) dx dy = 1$.

7.1 Independence, Conditional Probability, Expectation

7.1.1 Independence and conditional probability

Two random variables X and Y are **independent** if and only if $\forall B_X, B_Y \subseteq \mathbb{R}.$,

$$\mathrm{P}_{XY}(B_X, B_Y) = \mathrm{P}_X(B_X) \mathrm{P}_Y(B_Y).$$

More specifically,

Definition 7.1.1. *Two continuous random variables X and Y are* **independent** *if and only if*

$$f_{XY}(x,y) = f_X(x) f_Y(y), \quad \forall x, y \in \mathbb{R}.$$

Definition 7.1.2. *For two random variables* X, Y *we define the* **conditional probability distribution** $\mathrm{P}_{Y|X}$ *by*

$$\mathrm{P}_{Y|X}(B_Y \mid B_X) = \frac{\mathrm{P}_{XY}(B_X, B_Y)}{\mathrm{P}_X(B_X)}, \quad B_X, B_Y \subseteq \mathbb{R}.$$

This is the revised probability of Y falling inside B_Y *given* that we now know $X \in B_X$.

Then we have X and Y are independent $\iff \mathrm{P}_{Y|X}(B_Y \mid B_X) = \mathrm{P}_Y(B_Y),\ \forall B_X, B_Y \subseteq \mathbb{R}$.

Definition 7.1.3. *For random variables* X, Y *we define the* **conditional probability density function** $f_{Y|X}$ *by*

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x,y)}{f_X(x)}, \quad x, y \in \mathbb{R}.$$

Note The random variables X and Y are independent $\iff f_{Y|X}(y \mid x) = f_Y(y), \forall x, y \in \mathbb{R}$.

7.1.2 Expectation

Suppose we have a (measurable) bivariate function of interest of the random variables X and Y, i.e. $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$.

Definition 7.1.4. *If* X *and* Y *are discrete, we define* $\mathrm{E}\,(g(X,Y))$ *by*

$$\mathrm{E}\,(g(X,Y)) = \sum_{y \in \mathbb{Y}} \sum_{x \in \mathbb{X}} g(x,y) p_{XY}(x,y).$$

Definition 7.1.5. *If* X *and* Y *are jointly continuous, we define* $\mathrm{E}\,(g(X,Y))$ *by*

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

Immediately from these definitions we have the following:

- Expectation is always linear:

$$\mathrm{E}_{XY}[g_1(X,Y) + g_2(X,Y)] = E_{XY}[g_1(X,Y)] + E_{XY}[g_2(X,Y)].$$

- If $g(X,Y) = g_1(X) + g_2(Y)$,

$$\mathrm{E}_{XY}[g_1(X) + g_2(Y)] = E_X[g_1(X)] + E_Y[g_2(Y)].$$

- If $g(X,Y) = g_1(X)g_2(Y)$ and X and Y are **independent**,

$$\mathrm{E}_{XY}[g_1(X)g_2(Y)] = \mathrm{E}_X[g_1(X)]\,\mathrm{E}_Y[g_2(Y)].$$

  In particular, considering $g(X,Y) = XY$ for independent X, Y we have

$$\mathrm{E}_{XY}(XY) = \mathrm{E}_X(X)\,\mathrm{E}_Y(Y).$$

  Warning! In general $\mathrm{E}_{XY}(XY) \neq \mathrm{E}_X(X)\,\mathrm{E}_Y(Y)$.

7.1.3 Conditional Expectation

Suppose X and Y are discrete random variables with joint pmf $p(x,y)$. If we are given the value x of the random variable X, our revised pmf for Y is the conditional pmf $p(y|x)$, for $y \in \mathbb{R}$.

Definition 7.1.6. *The* **conditional expectation** *of Y given $X = x$ is*

$$E_{Y|X}(Y \mid X = x) = \sum_{y \in \mathbb{Y}} y\ p(y \mid x). \qquad \text{(discrete)}$$

$$E_{Y|X}(Y \mid X = x) = \int_{y=-\infty}^{\infty} y\ f(y \mid x) dy. \quad \text{(continuous)}$$

In either case, the conditional expectation is a function of x but not y.

For a single variable X we considered the expectation of $g(X) = (X - \mu_X)(X - \mu_X)$, called the variance and denoted σ_X^2.

The bivariate extension of this is the expectation of $g(X,Y) = (X - \mu_X)(Y - \mu_Y)$. We define the **covariance** of X and Y by

$$\sigma_{XY} = \mathrm{Cov}(X,Y) = \mathrm{E}_{XY}[(X - \mu_X)(Y - \mu_Y)].$$

Covariance measures how the random variables move in tandem with one another, and so is closely related to the idea of correlation.

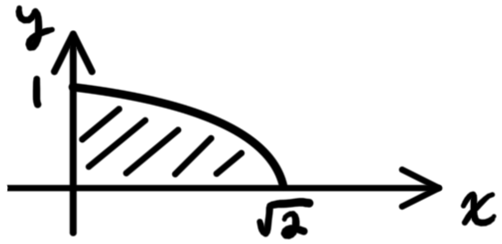
Definition 7.1.7. *We define the* **correlation** *of X and Y by*

$$\rho_{XY} = \mathrm{Cor}(X,Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Unlike the covariance, the correlation is invariant to the scale of the random variables X and Y.

It is easily shown that if X and Y are independent random variables, then $\sigma_{XY} = \rho_{XY} = 0$.

7.2 Examples



Example $f(x,y) = cxy$, $0 < x < \sqrt{2}$, $0 < y < \sqrt{1 - x^2/2}$ for some constant $c \in \mathbb{R}$, and $f(x,y) = 0$ outside this region.

Question Find c.

use normalization condition: $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$

$$1 = \int_0^{\sqrt{2}} dx \int_0^{\sqrt{1-x^2/2}} dy\; cxy$$

$$= c\int_0^{\sqrt{2}} dx\, x \tfrac{1}{2} y^2 \Big|_0^{\sqrt{1-x^2/2}} = \frac{c}{2}\int_0^{\sqrt{2}} dx\; x\left(1 - \frac{x^2}{2}\right)$$

$$= \frac{c}{2}\left[\frac{1}{2}(\sqrt{2})^2 - \frac{1}{8}(\sqrt{2})^4\right] = \frac{c}{4} \qquad \Rightarrow c = 4$$

Question Find $f_X(x)$.

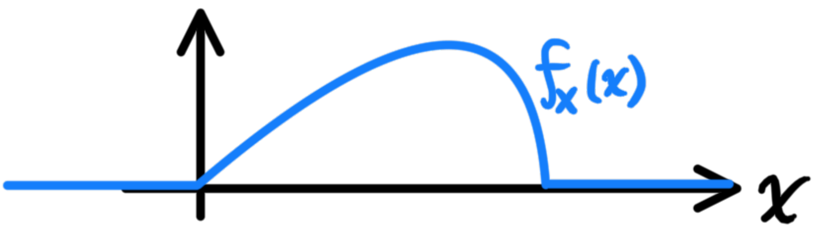
marginal pdf $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$

$$f_X(x) = \int_0^{\sqrt{1-x^2/2}} 4xy\,dy$$

$$= 4x\int_0^{\sqrt{1-x^2/2}} y\,dy$$

$$= 2x\left(1 - \frac{x^2}{2}\right) \quad \text{for } 0 < x < \sqrt{2}$$

and 0 otherwise



■

Example Suppose that the lifetime, X, and brightness, Y of a light bulb are modelled as continuous random variables. Let their joint pdf be given by

$$f(x,y) = \lambda_1\lambda_2 e^{-\lambda_1 x - \lambda_2 y}, \quad x, y > 0.$$

Question Are lifetime and brightness independent?

Solution If the lifetime and brightness are independent we would have

$$f(x,y) = f(x)f(y) \quad \text{for all } x, y \in \mathbb{R}.$$

The marginal pdf for X is

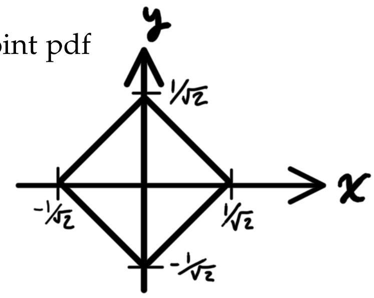
$$\begin{aligned} f(x) &= \int_{y=-\infty}^{\infty} f(x,y)dy = \int_{y=0}^{\infty} \lambda_1\lambda_2 e^{-\lambda_1 x - \lambda_2 y} dy \\ &= \lambda_1 e^{-\lambda_1 x}. \end{aligned}$$

Similarly $f(y) = \lambda_2 e^{-\lambda_2 y}$. Hence $f(x,y) = f(x)f(y)$ and X and Y are independent.

■

Example Suppose continuous random variables $(X, Y) \in \mathbb{R}^2$ have joint pdf

$$f(x,y) = \begin{cases} 1, & |x| + |y| < 1/\sqrt{2} \\ 0, & \text{otherwise.} \end{cases}$$



Question Determine the marginal pdfs for X and Y.

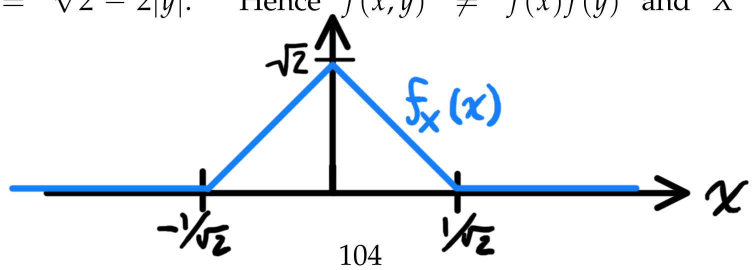
Solution

We have $|x| + |y| < 1/\sqrt{2} \iff |y| < 1/\sqrt{2} - |x|$. So

$$f(x) = \int_{y=-(\frac{1}{\sqrt{2}}-|x|)}^{\frac{1}{\sqrt{2}}-|x|} dy = \sqrt{2} - 2|x|.$$

For $|x| < \frac{1}{\sqrt{2}}$

Similarly $f(y) = \sqrt{2} - 2|y|$. Hence $f(x,y) \neq f(x)f(y)$ and X and Y are not independent.



■

7.3 Multivariate Transformations

7.3.1 Convolutions (sums of random variables)

Example Suppose X and Y are independent random variables with marginal distributions $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\eta)$. Let $Z = X + Y$ and find the pmf of Z.

For any $z = 0, 1, 2 \ldots$,

$$
\begin{aligned}
P(Z = z) &= \sum_{i=0}^{z} P(X = i \text{ and } Y = z - i) \\
&= \sum_{i=0}^{z} \frac{e^{-\lambda}\lambda^i}{i!} \frac{e^{-\eta}\eta^{z-i}}{(z-i)!} && \text{(X and Y are independent Poisson RVs)} \\
&= \frac{e^{-(\lambda+\eta)}}{z!} \sum_{i=0}^{z} \frac{z!}{i!(z-i)!} \lambda^i \eta^{z-i} && \text{(multiply and divide by $z!$ to get a binomial coefficient)} \\
&= \frac{e^{-(\lambda+\eta)}}{z!} (\lambda+\eta)^z && \text{(using binomial theorem)}
\end{aligned}
$$

$\Rightarrow Z \sim \text{Poisson}(\lambda + \eta)$.

Intuitive derivation

During Δt, $P(\bullet) = \frac{\lambda}{T}\Delta t$

During Δt, $P(\bullet) = \frac{\eta}{T}\Delta t$

During Δt, $P(\bullet \text{ or } \bullet) = \frac{\lambda+\eta}{T}\Delta t + \mathcal{O}(\Delta t^2)$

$\Downarrow$ Total # in time T $\sim \text{Poisson}(\lambda+\eta)$ ■

Next we can compute the conditional distribution of X given $Z = z$:

$$
\begin{aligned}
P(X = x \mid Z = z) \quad \propto \quad P(X = x \text{ and } Z = z) \quad &= \quad P(Y = z - x \text{ and } X = x) \\
&= \quad P(Y = z - x)P(X = x).
\end{aligned}
$$

[*Question* Given $Z = z$, what is the support of X? *Answer*: $\{0, 1, 2, \ldots, z\}$.]

$$
\begin{aligned}
p_{X|Z}(x|z) &\propto \frac{e^{-\eta}\eta^{z-x}}{(z-x)!} \frac{e^{-\lambda}\lambda^x}{x!} \\
&\propto \frac{1}{x!(z-x)!} \lambda^x \eta^{z-x} && \text{(ignoring factors independent of x)} \\
&\propto \frac{z!}{x!(z-x)!} \left(\frac{\lambda}{\lambda+\eta}\right)^x \left(\frac{\eta}{\lambda+\eta}\right)^{z-x} && \text{(introducing the factor $z!/(\lambda+\eta)^z$, which is independent of x)} \\
\Rightarrow X|Z = z \quad &\sim \quad \text{Binomial}\left(z, \frac{\lambda}{\lambda+\eta}\right).
\end{aligned}
$$

Theorem 7.4. *(Convolution Theorem) If X and Y are independent random variables and $Z = X + Y$, then*

$$p_Z(z) = \begin{cases} \sum_{x \in \mathbb{X}} p_X(x) p_Y(z-x) & \text{(discrete case)}, \\ \\ \int_{\mathbb{R}} f_X(\omega) f_Y(z-\omega)\, \mathrm{d}\omega & \text{(continuous case)}. \end{cases}$$

Note In the discrete case, we need only sum over all possible values of X such that Y can take on the value $z - x$.

Example Suppose $X \sim \mathrm{N}(0, \sigma^2)$ and $Y \sim \mathrm{N}(0, 1)$ with X and Y independent. Let $Z = X + Y$ and derive the pdf of Z.

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x^2}{2\sigma^2}\right] \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(z-x)^2}{2}\right] \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma^2} + (z-x)^2\right)\right] \mathrm{d}x \end{aligned}$$

The trick is completing the square,

$$\begin{aligned} \left(\frac{x^2}{\sigma^2} + (z-x)^2\right) &= \left(1 + \frac{1}{\sigma^2}\right) x^2 - 2zx + z^2 \\ &= c^2 \left(x^2 - \frac{2z}{c^2} x\right) + z^2 \qquad \left(\text{setting } c^2 = 1 + \frac{1}{\sigma^2}\right) \\ &= c^2 \left(x - \frac{z}{c^2}\right)^2 - \frac{z^2}{c^2} + z^2 \end{aligned}$$

So,

$$\begin{aligned} f_Z(z) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} c^2 \left(x - \frac{z}{c^2}\right)^2\right] \exp\left[-\frac{1}{2}\left(1 - \frac{1}{c^2}\right) z^2\right] \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi\sigma^2 c^2}} \exp\left[-\frac{z^2}{2(1+\sigma^2)}\right] \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1/c^2)}} \exp\left[-\frac{1}{2} c^2 \left(x - \frac{z}{c^2}\right)^2\right] \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi(1+\sigma^2)}} \exp\left[-\frac{z^2}{2(1+\sigma^2)}\right] \qquad \text{(the integrand is the pdf of a } \mathrm{N}(z/c^2, 1/c^2)\text{, so it integrates to 1)} \end{aligned}$$

$$\Rightarrow Z \sim \mathrm{N}(0, 1 + \sigma^2).$$

■

Theorem 7.5. *If* $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$ *and* $Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$ *with* X *and* Y *independent, then* $Z = X + Y \sim \mathrm{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Note that

$$Z = X + Y = \sigma_Y \left[\underbrace{\frac{X - \mu_X}{\sigma_Y}}_{\sim N\left(0, \frac{\sigma_X^2}{\sigma_Y^2}\right)} + \underbrace{\frac{Y - \mu_Y}{\sigma_Y}}_{\sim N(0,1)} \right] + \mu_X + \mu_Y = \sigma_Y W + \mu_X + \mu_Y \tag{7.1}$$

Setting W equal to the term in brackets, $W = \left(\frac{X-\mu_X}{\sigma_Y} + \frac{Y-\mu_Y}{\sigma_Y} \right)$, by the above example we have

$$W \sim \mathrm{N}\left(0, 1 + \frac{\sigma_X^2}{\sigma_Y^2}\right) \text{ and thus } Z \sim \mathrm{N}\left(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2\right).$$

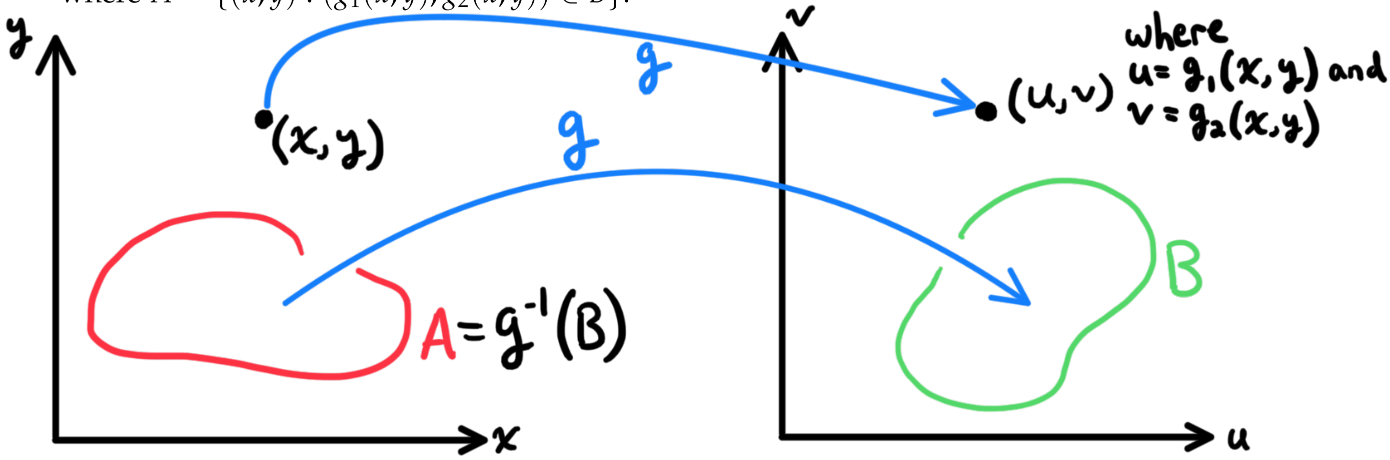
[We used the fact that $Z = aW + b$ for some constants a and b.
In particular, from (7.1), $\mathrm{E}(Z) = \mu_X + \mu_Y$ and $\mathrm{Var}(Z) = \sigma_X^2 + \sigma_Y^2$.]

7.5.1 General Bivariate Transformations

Suppose (X,Y) is a bivariate random variable and let $U = g_1(X,Y)$ and $V = g_2(X,Y)$. Then for any $B \subseteq \mathbb{R}^2$,

$$\mathrm{P}\Big((U,V) \in B\Big) = \mathrm{P}\Big((X,Y) \in A\Big),$$

where $A = \{(x,y) : (g_1(x,y), g_2(x,y)) \in B\}$.

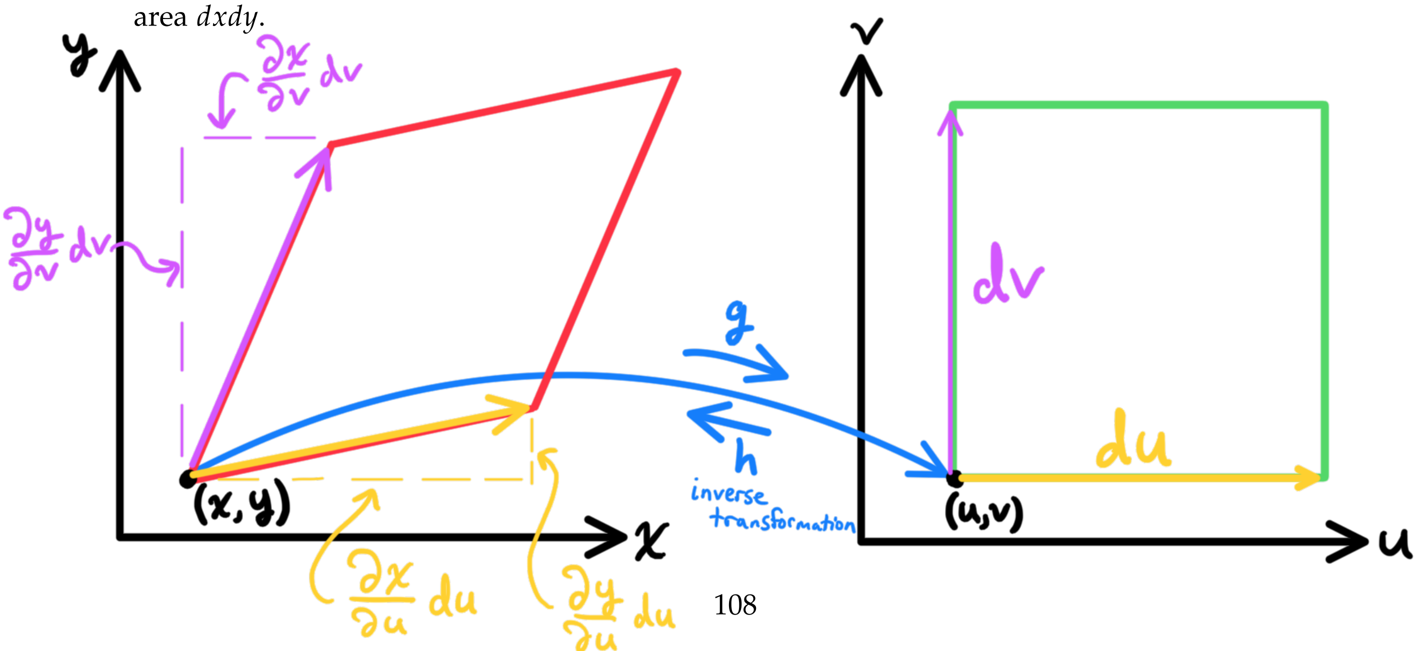


Case 1: If (X,Y) is discrete: Let $A(u,v) = \{(x,y) \in (\mathbb{X},\mathbb{Y}) : (g_1(x,y), g_2(x,y)) = (u,v)\}$, then

$$p_{UV}(u,v) = \mathrm{P}(U=u, V=v) = \mathrm{P}\Big((X,Y) \in A(u,v)\Big) = \sum_{\substack{(x,y):\\ g_1(x,y)=u\\ g_2(x,y)=v}} p_{XY}(x,y).$$

Case 2: If (X,Y) is continuous: We will assume $\mathbf{g} = (g_1, g_2)$ defines a one-to-one transformation between (x,y) and (u,v).

The most direct route to the transformation law for the joint pdf is to imagine a tiny region with area $dudv$ centered at the point (u,v) in the (u,v)-plane. The probability that the joint RV (U,V) will be in this region is $f_{UV}(u,v)dudv$ (in the limit that du and dv become small). This tiny region is the image, under $\mathbf{g}$, of a corresponding tiny region in the (x,y)-plane with area $dxdy$.



The area $dxdy$ is related to the area $dudv$ by the absolute value of the **Jacobian determinant**: $dxdy = |J|dudv$, where

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

and x, y, u, and v are related to each other by the transformation functions g_1 and g_2.

Thus, the probability that the joint RV (U, V) is in the region with area $dudv$ is the same as the probability that (X, Y) is in the region with area $dxdy$, i.e.

$$f_{UV}(u,v)dudv = f_{XY}(x,y)dxdy = f_{XY}(x,y)\left|\frac{\partial(x,y)}{\partial(u,v)}\right| dudv,$$

or

$$f_{UV}(u,v) = f_{XY}(x,y)\left|\frac{\partial(x,y)}{\partial(u,v)}\right|.$$

Remember absolute value

An equivalent way to extract the transformation law for the joint pdf is by changing coordinates in a multidimensional integral. We will assume (g_1, g_2) defines a one-to-one transformation between (x, y) and (u, v). Let h_1 and h_2 be the inverse function, such that $x = h_1(u, v)$ and $y = h_2(u, v)$. Perform a change of coordinates in the integral on the second line below:

$$\begin{aligned}
\int_B f_{UV}(u,v)dudv = \mathrm{P}\Big((U,V) \in B\Big) &= \mathrm{P}\Big((X,Y) \in \mathbf{g}^{-1}(B)\Big) \\
&= \int_{\mathbf{g}^{-1}(B)} f_{XY}(x,y)dxdy \\
&= \int_B f_{XY}\Big(h_1(u,v), h_2(u,v)\Big)|J|\,dudv,
\end{aligned}$$

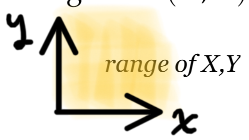
where $|J|$ is the absolute value of the Jacobian determinant,

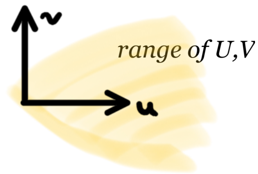
$$|J| = \begin{vmatrix} \dfrac{\partial}{\partial u}h_1(u,v) & \dfrac{\partial}{\partial v}h_1(u,v) \\ \\ \dfrac{\partial}{\partial u}h_2(u,v) & \dfrac{\partial}{\partial v}h_2(u,v) \end{vmatrix} = \left|\frac{\partial}{\partial u}h_1(u,v)\cdot\frac{\partial}{\partial v}h_2(u,v) - \frac{\partial}{\partial v}h_1(u,v)\cdot\frac{\partial}{\partial u}h_2(u,v)\right|.$$

Since this holds for arbitrary an region B we have,

$$\boxed{f_{UV}(u,v) = f_{XY}\Big(h_1(u,v), h_2(u,v)\Big)|J|.}$$

Warning: Do not forget the range of the joint random variables (U, V). This can be found from the transformation $\mathbf{g}$ and the range of (X, Y). A simple range for (X, Y) (e.g. $\mathbb{X} = \mathbb{Y} = \mathbb{R}^+$) very often corresponds with a non-trivial range for (U, V) if the transformation law is nonlinear.





Example (Sum of Gamma Distributions)

The *Gamma function* is defined by:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt, \text{ for } \alpha > 0.$$

A few important properties:

(1) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$, for $\alpha > 1$.
(2) $\Gamma(1) = \int_0^\infty e^{-t}\, dt = 1$.
(3) $\Gamma(n) = (n-1)!$, for any integer n.
(4) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Now, by the definition of the gamma function,

$$f_Y(y) = \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y}, \text{ for } y > 0 \text{ and } \alpha > 0,$$

is a valid pdf. Now, let $X = Y/\beta$. The pdf of X is

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-x\beta}, \text{ where } 0 < x < +\infty \text{ and } \alpha, \beta > 0. \tag{7.2}$$

A continuous random variable with the pdf given in (7.2) is said to follow the *gamma distribution*. Note that, like the exponential distribution, the Gamma distribution has two standard formulations:

1. $f_X(x) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$, for $0 < x < +\infty$ and $\alpha, \beta > 0$;
2. $f_X(x) = \dfrac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}$, for $0 < x < +\infty$ and $\alpha, \theta > 0$.

The first form has $\mathrm{E}(X) = \alpha/\beta$ and $\mathrm{Var}(X) = \alpha/\beta^2$, and the second form has $\mathrm{E}(X) = \alpha\theta$ and $\mathrm{Var}(X) = \alpha\theta^2$.

α, β α, β

Now, suppose $X \sim \text{Gamma}(\lambda, 1)$ and $Y \sim \text{Gamma}(\xi, 1)$ with X and Y independent. Let $Z_1 = X + Y$ and $Z_2 = \dfrac{X}{X+Y}$ and find the joint pdf of Z_1 and Z_2 and both of the marginal distributions.

Question Is $(X, Y) \to (Z_1, Z_2)$ invertible? That is, is this a one-to-one transformation?

Yes. Notice that $Z_2 = \dfrac{X}{Z_1}$, so $X = Z_1 Z_2$ and $Y = Z_1 - X = Z_1(1 - Z_2)$. This defines the inverse transformation $x = h_1(z_1, z_2) = z_1 z_2$ and $y = h_2(z_1, z_2) = z_1(1 - z_2)$.

Since the support of (X, Y) is $\mathbb{R}^+ \times \mathbb{R}^+$ the range of (Z_1, Z_2) is $\mathbb{R}^+ \times (0, 1)$.

Compute the Jacobian of the transformation:

$$|J| = \begin{vmatrix} \dfrac{\partial x}{\partial z_1} & \dfrac{\partial x}{\partial z_2} \\ \dfrac{\partial y}{\partial z_1} & \dfrac{\partial y}{\partial z_2} \end{vmatrix} = \begin{vmatrix} z_2 & z_1 \\ 1 - z_2 & -z_1 \end{vmatrix} = \left| -z_1 z_2 - z_1(1 - z_2) \right|$$

$$= |-z_1| = z_1 .$$

Starting with

$$f_{XY}(x, y) = \frac{1}{\Gamma(\lambda)\Gamma(\xi)}\, x^{\lambda-1}\, y^{\xi-1}\, e^{-(x+y)} \quad \text{for} \quad x, y > 0,$$

write x and y in terms of z_1 and z_2

find $f_{Z_1,Z_2}(z_1, z_2) = f_{XY}(x, y) \left| \dfrac{\partial(x, y)}{\partial(z_1, z_2)} \right|$ ~ z_1

$$= \frac{1}{\Gamma(\lambda)\Gamma(\xi)} (z_1 z_2)^{\lambda-1} \left[z_1(1 - z_2) \right]^{\xi-1} e^{-z_1}\, z_1$$

$$= \underbrace{\frac{1}{\Gamma(\lambda+\xi)} z_1^{\lambda+\xi-1} e^{-z_1}}_{\text{pdf of Gamma}(\lambda+\xi,\,1)} \ \underbrace{\frac{\Gamma(\lambda+\xi)}{\Gamma(\lambda)\Gamma(\xi)} z_2^{\lambda-1} (1 - z_2)^{\xi-1}}_{Z_2 \sim \text{Beta}(\lambda,\,\xi)}$$

$\Rightarrow Z_1 \sim \text{Gamma}(\lambda+\xi, 1)$

and Z_1, Z_2 are independent.

■

Note, a random variable, X, having pdf

$$f_X(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} \text{ for } 0 < x < 1,$$

is said to follow a *beta distribution* and we write $X \sim \text{Beta}(\alpha, \beta)$. Here α and β are two positive parameters and

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

is called the *beta function*.

Theorem 7.6. *If $X \sim \text{Gamma}(\lambda, \theta)$ and $Y \sim \text{Gamma}(\xi, \theta)$ with X and Y independent, then $Z = X + Y \sim \text{Gamma}(\lambda + \xi, \theta)$.*

Proof: Option 1) use the same method as in the above example with the normal distribution. Option 2) find the joint pdf of $Z_1 = X + Y$ and $Z_2 = X/(X+Y)$ and then integrate it over z_2 to obtain the marginal pdf of Z_1.