Appendix

Results for the Gamma distribution

The Gamma distribution arises in several contexts:

- as the continuous analogue of the Negative Binomial distribution
- as a function related to the Poisson cdf
- as a consequence of the Poisson process.

Connection to the Geometric distribution

Recall the origin of the Poisson distribution; we consider

$$Y \sim Binomial(n, p)$$

then fix $\lambda = np$, and allow $n \longrightarrow \infty$.

- We have an infinite sequence of independent Bernoulli (binary) trials, where the successes occur at a constant rate of λ, independently of the results of other trials.
- Furthermore, the number of trials between successes are independent $Geometric(p)$ random variables

$$0000100010010101000011\cdots$$

Connection to the Geometric distribution (cont.)

In the sequence of infinite independent Bernoulli trials, let Y_1 be the number of trials until the first success. Then

$$Y_1 \sim \mathit{Geometric}(p),$$

with

$$p_{Y_1}(y) = (1-p)^{y-1}p \qquad y = 1, 2, 3, \ldots$$

with $p_{Y_1}(y)$ otherwise, with corresponding cdf

$$F_{Y_1}(y) = \begin{cases} 0 & y < 1 \\ 1 - (1-p)^{\lfloor y \rfloor} & y \geq 1 \end{cases}$$

where $\lfloor y \rfloor$ is the largest integer no greater than y.

Connection to the Geometric distribution (cont.)

Now, let $X_1 = Y_1/n$, and $\lambda = np$. Then for fixed x

$$F_{X_1}(x) = P(X_1 \leq x) = P(Y_1 \leq nx) = F_{Y_1}(nx)$$

that is

$$F_{X_1}(x) = \begin{cases} 0 & x < 1/n \\ 1 - \left(1 - \frac{\lambda}{n}\right)^{\lfloor nx \rfloor} & x \geq 1/n \end{cases}$$

Connection to the Geometric distribution (cont.)

As $n \longrightarrow \infty$

$$\left(1 - \frac{\lambda}{n}\right)^{\lfloor nx \rfloor} \longrightarrow \exp\{-\lambda x\}$$

for $x > 0$. That is

$$F_{X_1}(x) \longrightarrow \begin{cases} 0 & x < 0 \\ 1 - \exp\{-\lambda x\} & x \geq 0 \end{cases}$$

Thus $X_1 \sim \mathit{Exponential}(1/\lambda)$.

Connection to the Geometric distribution (cont.)

This result also follows using mgfs: we have that

$$m_{Y_1}(t) = \frac{pe^t}{1-qe^t} \quad \Longrightarrow \quad m_{X_1}(t) = m_{Y_1}\left(\frac{t}{n}\right) = \frac{pe^{t/n}}{1-qe^{t/n}}.$$

If $p = \lambda/n$, then

$$m_{X_1}(t) = \frac{(\lambda/n)e^{t/n}}{1-(1-\lambda/n)e^{t/n}} = \frac{\lambda e^{t/n}}{\lambda e^{t/n} + n(1-e^{t/n})} \longrightarrow \frac{\lambda}{\lambda - t}$$

as $n \longrightarrow \infty$, as $e^{t/n} \longrightarrow 1$ and

$$n(1-e^{t/n}) = n\left(1 - 1 - \frac{t}{n} - \frac{t^2}{2n^2} - \cdots\right) \longrightarrow -t.$$

Connection to the Geometric distribution (cont.)

Suppose $X_1, X_2, X_3, \ldots, X_r, \ldots$ are defined in terms of the infinite sequence of independent $Bernoulli(p)$ trials and the random variables

$$Y_r = \text{“number of trials between } (r-1)\text{th and } r\text{th success”}$$

via the relationship

$$X_r = \frac{Y_r}{n}.$$

Then in the limit as $n \longrightarrow \infty$ with $\lambda = np$ fixed

$$X_1, X_2, X_3, \ldots, X_r, \ldots \sim Exponential(1/\lambda)$$

with these variables independent.

Connection to the Poisson distribution

Let $Y \sim Gamma(r, 1/\lambda)$ for positive integer $r = 1, 2, \ldots$ and consider for $y \geq 0$ the cdf of Y,

$$F_Y(y) = P(Y \leq y) = \int_0^y \frac{\lambda^r}{(r-1)!} u^{r-1} e^{-\lambda u} \, du$$

with $F_Y(y) = 0$ for $y < 0$. Define

$$S_Y(r, y) = P(Y > y) = \int_y^\infty \frac{\lambda^r}{(r-1)!} u^{r-1} e^{-\lambda u} \, du$$

for $y > 0$.

Connection to the Poisson distribution (cont.)

Integrating by parts, we have that

$$S_Y(r,y) = \left[-\frac{\lambda^{r-1}}{(r-1)!}u^{r-1}e^{-\lambda u}\right]_y^\infty + \int_y^\infty \frac{\lambda^{r-1}}{(r-2)!}u^{r-2}e^{-\lambda u}\,du$$

$$= \frac{\lambda^{r-1}}{(r-1)!}y^{r-1}e^{-\lambda y} + S_Y(r-1,y)$$

$$= \frac{\lambda^{r-1}}{(r-1)!}y^{r-1}e^{-\lambda y} + \frac{\lambda^{r-2}}{(r-2)!}y^{r-2}e^{-\lambda y} + S_Y(r-2,y)$$

using the recursion.

Connection to the Poisson distribution (cont.)

By extending the recursion, we find that

$$S_Y(r, y) = \sum_{j=1}^{r-1} \frac{\lambda^j}{j!} y^j e^{-\lambda y} + S_Y(1, y).$$

We also have by direct calculation that

$$S_Y(1, y) = \int_y^\infty \lambda e^{-\lambda u} \, du = e^{-\lambda y}$$

so

$$S_Y(r, y) = \sum_{j=0}^{r-1} \frac{\lambda^j}{j!} y^j e^{-\lambda y}.$$

Connection to the Poisson distribution (cont.)

Therefore, evaluating at $y = 1$

$$S_Y(r, 1) = \sum_{j=0}^{r-1} \frac{\lambda^j}{j!} e^{-\lambda}.$$

Thus we have that if $Y \sim Gamma(r, 1/\lambda)$,

$$P(Y > 1) = \sum_{j=0}^{r-1} \frac{\lambda^j}{j!} e^{-\lambda}.$$

Connection to the Poisson distribution (cont.)

Recall the *Poisson process* definition of the Poisson distribution.

- The process $X(t)$ defines a collection of discrete random variables whose probabilistic properties are determined as a limiting case of a sequence of independent Bernoulli random variables.
- Events occur in continuous time at a constant rate λ, with events occurring independently;
- The number of events that occur in any interval of length t is a random variable having a Poisson distribution with parameter λt.

Connection to the Poisson distribution (cont.)

The cdf of $X \equiv X(1)$ is therefore

$$F_X(x) = P(X \leq x) = \sum_{j=0}^{x} e^{-\lambda} \frac{\lambda^j}{j!}$$

for non-negative integer x. Hence

$$P(X \leq x-1) = \sum_{j=0}^{x-1} e^{-\lambda} \frac{\lambda^j}{j!} = P(Y > 1)$$

where $Y \sim Gamma(x, 1/\lambda)$.

Connection to the Negative Binomial distribution

The negative binomial distribution arises in the context of a sequence of independent binary trials with success probability p as the distribution of number of trials, Y_r, required to obtain the rth success.

The mgf of $Y_r \sim NegBinomial(r, p)$ takes the form

$$m_{Y_r}(t) = \left(\frac{pe^t}{1 - qe^t} \right)^r .$$

Connection to the Negative Binomial distribution

We now let the number of trials n tend to infinity, but fix $\lambda = np$. If we let $X_r = Y_r/n$ for fixed $r = 1, 2, 3, \ldots$, we have that

$$m_{X_r}(t) \longrightarrow m_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r$$

for $t < \lambda$. Thus the limiting mgf corresponds to the limiting random variable

$$X \sim Gamma(r, 1/\lambda).$$

Transformations

Discrete case: In the discrete case, we consider how the probability attached to individual values is transported under the transformation.

Suppose Y takes values on the (countable) set

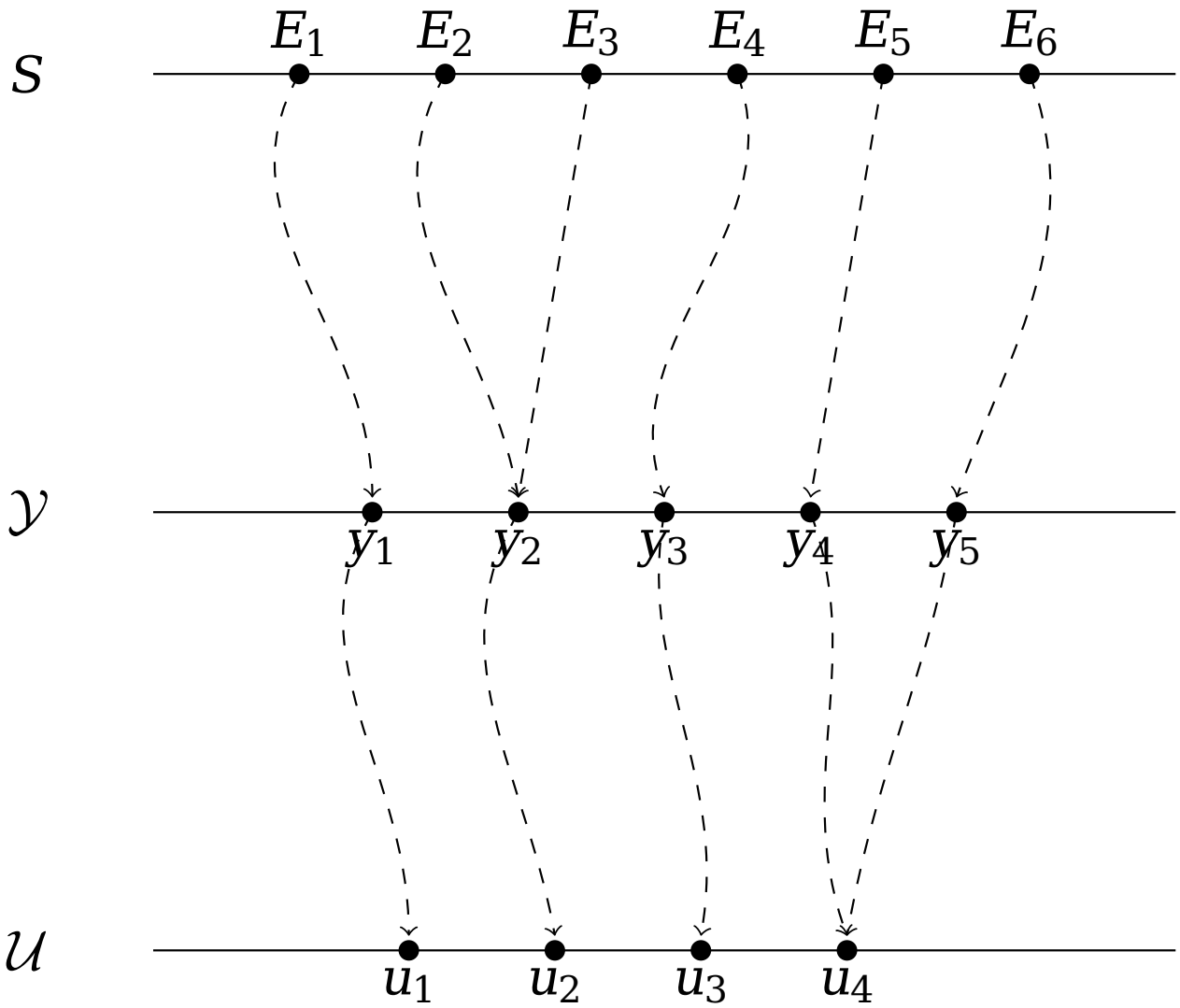
$$\{y_1, y_2, \ldots, y_n, \ldots\}.$$

Under the transformation $h(.)$, these values map to the set

$$\{u_1, u_2, \ldots, u_n, \ldots\}$$

where $u_i = h(y_j)$ for at least one j.

Transformations (cont.)



Transformations (cont.)

In the above example

$$u_1 = h(y_1) \quad u_2 = h(y_2) \quad u_3 = h(y_3) \quad u_4 = h(y_4) = h(y_5)$$

We typically deal with simple transformations, for example

$$U = 2Y \qquad U = \exp(-Y) \qquad U = \frac{Y}{Y+1} \qquad U = Y^2$$

Transformations (cont.)

To derive the probability distribution of U, we simply see how the probability accumulates on individual values under the transformation.

In the picture above, we have four values $\{u_1, u_2, u_3, u_4\}$ with probability as follows:

$$p_U(u) = P(U = u) = \begin{cases} p_Y(y_1) & u = u_1 \\ p_Y(y_2) & u = u_2 \\ p_Y(y_3) & u = u_3 \\ p_Y(y_4) + p_Y(y_5) & u = u_4 \end{cases}$$

Transformations (cont.)

Continuous case: The obstacle to using the same approach here is that for a continuous random variable Y we have

$$P(Y = y) = 0$$

for all real values y. To get around this, we can instead consider

$$P(Y \in \mathcal{A})$$

for a subset (eg interval) of the real line. For example, we could consider $\mathcal{A} = [2, 5]$, and the probability

$$P(Y \in \mathcal{A}) \equiv P(2 \leq Y \leq 5).$$

Transformations (cont.)

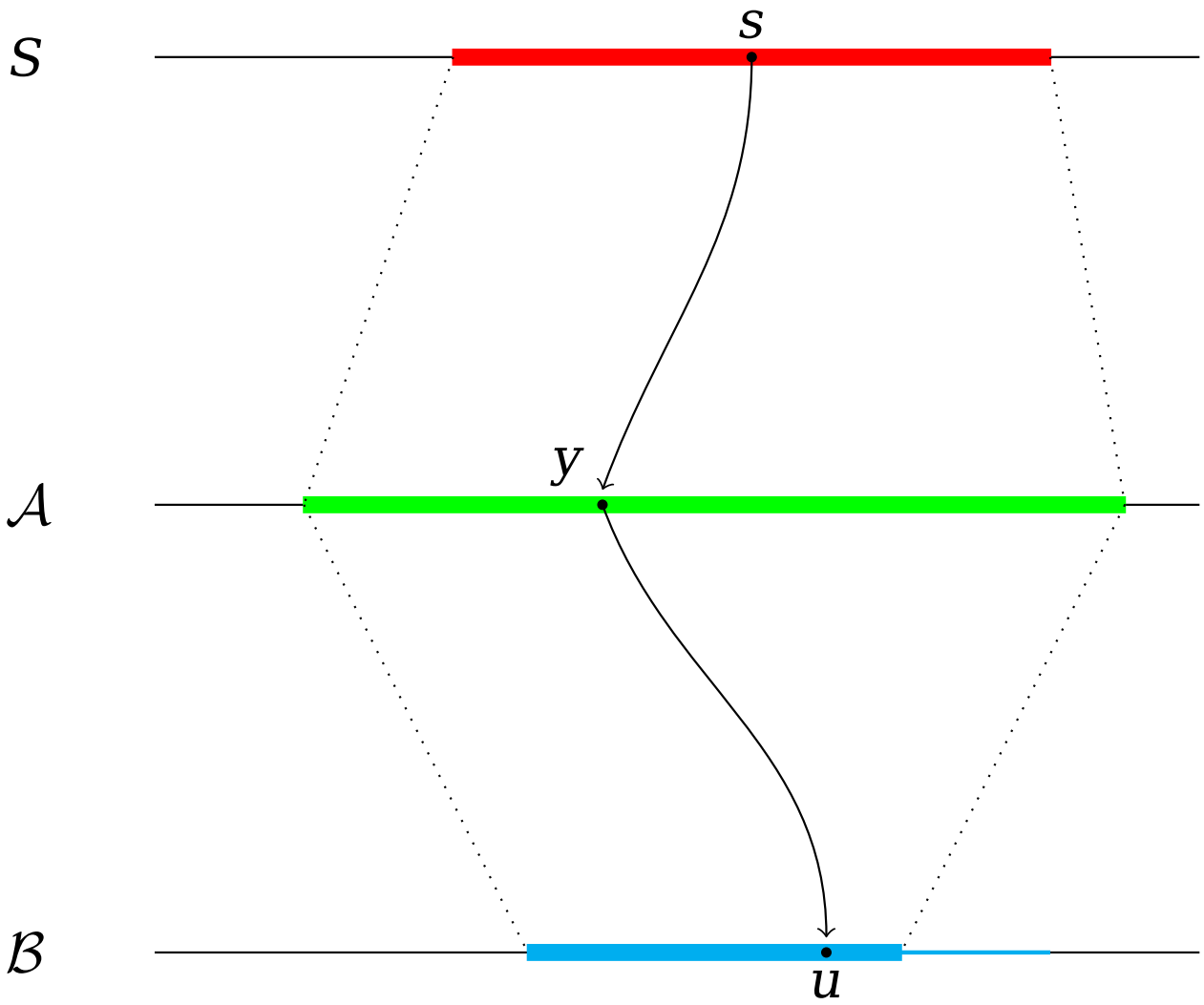
Suppose that $\mathcal{B}$ represents the set of values containing all the points that points in $\mathcal{A}$ can be mapped to under the transformation $h(.)$. That is,

$$\mathcal{B} = \{u : u = h(y), \text{some } y \in \mathcal{A}\}$$

Then we have that probability is preserved under the mapping

$$P(Y \in \mathcal{A}) = P(U \in \mathcal{B}).$$

Transformations (cont.)



Transformations (cont.)

Example

Suppose Y is a continuous random variable that can take values on $[0, \infty)$, and $U = h(Y) = 2Y$. This means that we need to consider the set

$$\mathcal{U} = \{u : u = 2y, \text{ some } y \in [0, \infty)\} \equiv [0, \infty)$$

that is, every point y gets mapped to $u = 2y$.

Transformations (cont.)

Example

Consider interval $\mathcal{A} = [a, b]$ for $a < b$; if we let $c = h(a) = 2a$ and $d = 2b$, then we have that

$$P(a \leq Y \leq b) = P(c \leq U \leq d).$$

In terms of the cdfs, we have that

$$F_Y(b) - F_Y(a) = F_U(d) - F_U(c)$$

and letting $a \longrightarrow 0$, we have that

$$F_Y(b) = P(Y \leq b) = P(U \leq d) = F_U(d).$$

Transformations (cont.)

Example

That is, for $b \in \mathbb{R}$.

$$F_Y(b) = F_U(2b)$$

Now, given F_Y, we wish to compute F_U. We can do this by noting that, for arbitrary u

$$\begin{aligned} F_U(u) &= P(U \leq u) = P(U/2 \leq u/2) \\ &= P(Y \leq u/2) \\ &= F_Y(u/2). \end{aligned}$$

Transformations (cont.)

Example

To get to the pdf, we need to differentiate both sides with respect to u:

$$f_U(u) = \frac{dF_U(u)}{du} = \frac{dF_Y(u/2)}{du} = \frac{1}{2} f_Y(u/2)$$

by the chain rule. For example, if $F_Y(y) = 1 - \exp\{-y\}$, then for $u \geq 0$,

$$f_U(u) = \frac{d}{du} \{1 - \exp\{-u/2\}\} = \frac{1}{2} \exp\{-u/2\}.$$

Transformations (cont.)

If $h(.)$ is not 1–1 (non-monotonic), then the identity

$$P(U \in \mathcal{B}) = P(Y \in h^{-1}(\mathcal{B}))$$

still holds, we just need more care in computing the right-hand side.

Transformations (cont.)

Example

Suppose that Y has a continuous uniform distribution on the range $[0, 1]$, and that

$$U = Y(1 - Y)$$

that is $U = h(Y)$ with $h(t) = t(1 - t)$.

First, as $\mathcal{Y} = [0, 1]$, we have that

$$\mathcal{U} = [0, 1/4].$$

Transformations (cont.)

Example

Then for $0 \leq u \leq 1/4$, we have that

$$P(U \leq u) = P(Y(1-Y) \leq u) = P(Y - Y^2 - u \leq 0)$$

That is

$$F_U(u) = P(Y^2 - Y + u \geq 0) = P((Y - y_1)(Y - y_2) \geq 0)$$

where, factorizing the quadratic, we have

$$(y_1, y_2) = \left(\frac{1 - \sqrt{1-4u}}{2}, \frac{1 + \sqrt{1-4u}}{2} \right)$$

Transformations (cont.)

Example

Now, in the quadratic

$$(y - y_1)(y - y_2) \geq 0 \qquad \Longleftrightarrow y \leq y_1 \quad \text{or} \quad y \geq y_2$$

so for $0 \leq u \leq 1/4$

$$\begin{aligned} F_U(u) &= P((Y - y_1)(Y - y_2) \geq 0) \\ &= P(Y \leq y_1) + P(Y \geq y_2) \\ &= F_Y(y_1) + (1 - F_Y(y_2)) \\ &= y_1 + 1 - y_2 \\ &= 1 - \sqrt{1 - 4u}. \end{aligned}$$

Transformations (cont.)

Example

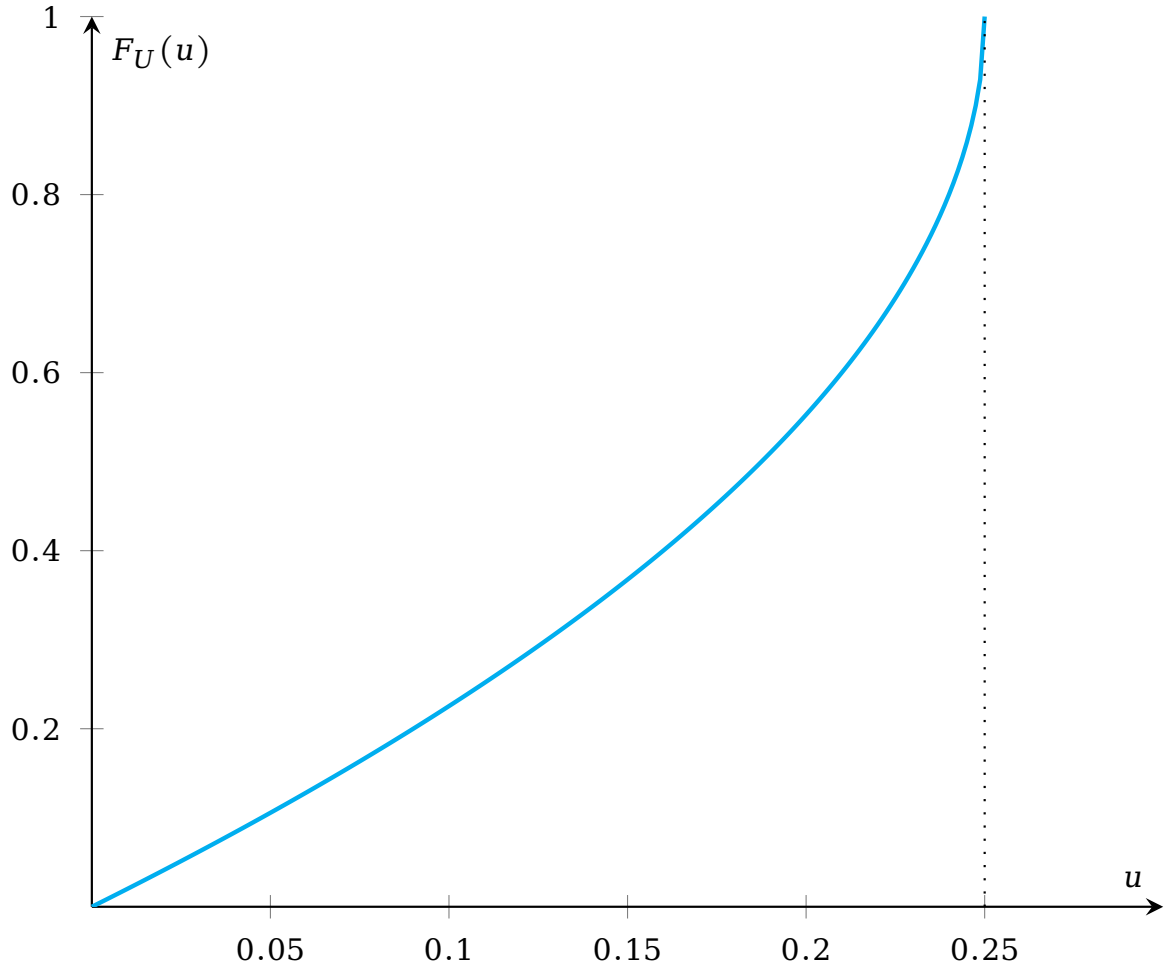
Differentiating, we have

$$f_U(u) = \frac{2}{\sqrt{1-4u}} \qquad 0 \leq u \leq 1/4.$$

For $u < 0$ and $u > 1/4$, $f_U(u) = 0$.

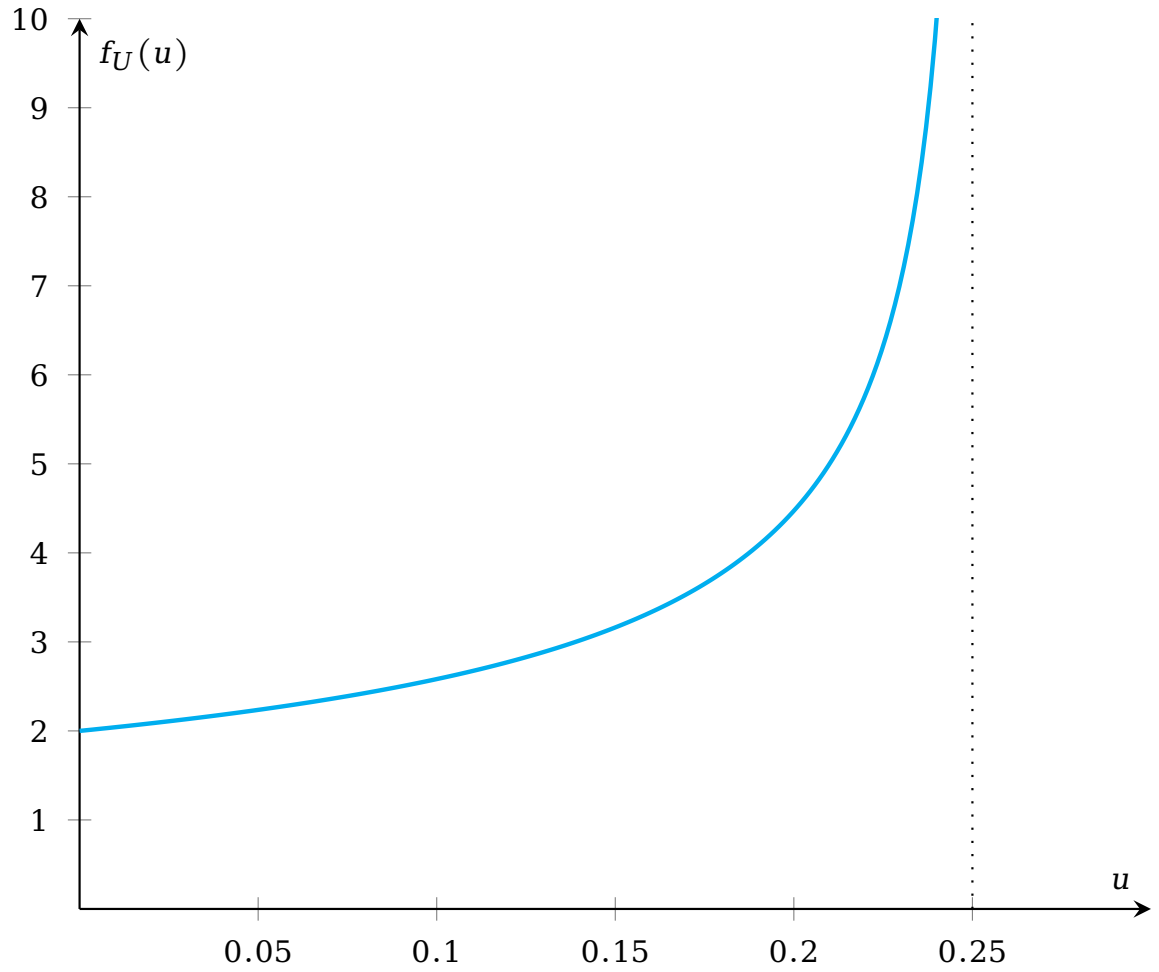
Transformations (cont.)

$$F_U(u) = 1 - \sqrt{1 - 4u}.$$



Transformations (cont.)

$$f_U(u) = 2/\sqrt{1-4u}.$$



Summations

Example (Poisson random variables)

If Y_1 and Y_2 are independent Poisson random variables with parameters λ_1 and λ_2, then if $Y = Y_1 + Y_2$,

$$Y \sim Poisson(\lambda_1 + \lambda_2)$$

Summations (cont.)

Example (Poisson random variables)

$$P(Y=y) = \sum_{y_1=0}^{y} P(Y_1=y_1)P(Y_2=y-y_1)$$

$$= \sum_{y_1=0}^{y} \frac{\lambda_1^{y_1} e^{-\lambda_1}}{y_1!} \frac{\lambda_2^{y-y_1} e^{-\lambda_2}}{(y-y_1)!}$$

$$= e^{-(\lambda_1+\lambda_2)} \lambda_2^{y} \sum_{y_1=0}^{y} \left(\frac{\lambda_1}{\lambda_2}\right)^{y_1} \frac{1}{y_1!(y-y_1)!}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)} \lambda_2^{y}}{y!} \left(1+\frac{\lambda_1}{\lambda_2}\right)^{y}$$

Summations (cont.)

Example (Poisson random variables)

But

$$\lambda_2^y \left(1 + \frac{\lambda_1}{\lambda_2}\right)^y = (\lambda_1 + \lambda_2)^y$$

so therefore

$$P(Y = y) = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1 + \lambda_2)^y}{y!} \qquad y = 0, 1, 2, \ldots$$

that is, Y has a Poisson distribution with parameter $\lambda_1 + \lambda_2$.

Summations (cont.)

Example (Exponential random variables)

Suppose

$$Y_1 : f_{Y_1}(y_1) = \frac{1}{\beta_1} e^{-y_1/\beta_1} \qquad y_1 \geq 0$$

$$Y_2 : f_{Y_2}(y_2) = \frac{1}{\beta_2} e^{-y_2/\beta_2} \qquad y_2 \geq 0$$

are independent random variables, and let $Y = Y_1 + Y_2$. For convenience, write

$$\lambda_1 = \frac{1}{\beta_1} \qquad \lambda_2 = \frac{1}{\beta_2}.$$

Example (Exponential random variables)

Then from the convolution formula, for $y \geq 0$,

$$
\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{Y_2}(y - y_1) f_{Y_1}(y_1)\, dy_1 \\
&= \int_0^y \lambda_2 e^{-\lambda_2 (y - y_1)} \lambda_1 e^{-\lambda_1 y_1}\, dy_1 \\
&= \lambda_1 \lambda_2 e^{-\lambda_2 y} \int_0^y e^{-(\lambda_1 - \lambda_2) y_1}\, dy_1
\end{aligned}
$$

Summations (cont.)

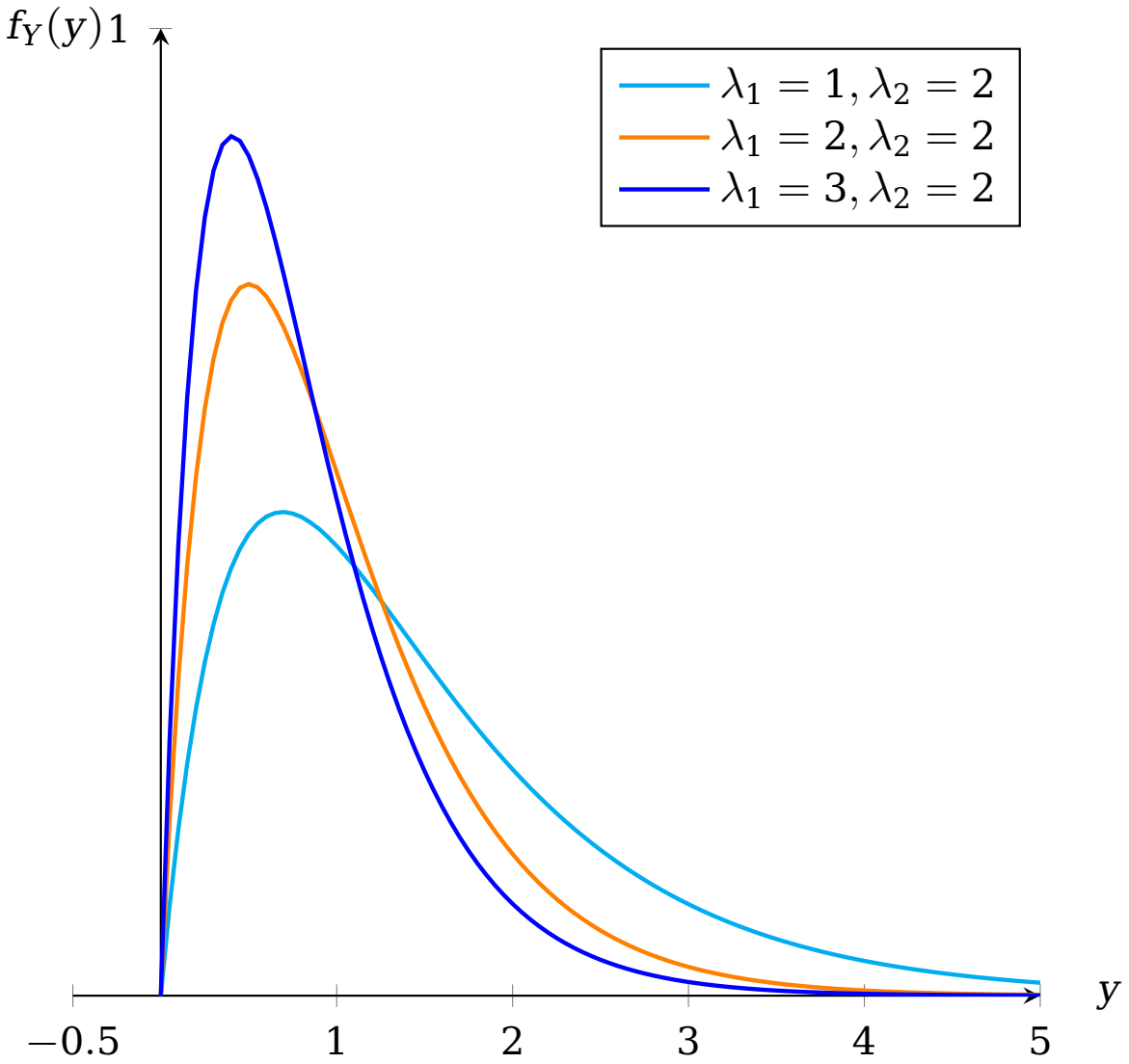
Example (Exponential random variables)

We then have two cases to consider:

$$f_Y(y) = \begin{cases} \dfrac{\lambda_1\lambda_2}{\lambda_1 - \lambda_2} e^{-\lambda_2 y}\left(1 - e^{-(\lambda_1-\lambda_2)y}\right) & \lambda_1 \neq \lambda_2 \\ \lambda^2 y e^{-\lambda y} & \lambda_1 = \lambda_2 = \lambda, \text{ say.} \end{cases}$$

In the second case, we see that in fact Y has a gamma distribution with parameters $\alpha = 2$ and $\beta = 1/\lambda$.

Summations (cont.)



Markov's inequality

Suppose that $g(Y)$ is a function of continuous random variable Y such that

$$\mathbb{E}[|g(Y)|] < \infty.$$

For any $k > 0$, let

$$\mathcal{A}_k = \{y \in \mathbb{R} : |g(y)| \geq k\}.$$

Markov's inequality (cont.)

Then

$$\begin{aligned}
\mathbb{E}[|g(Y)|] &= \int_{-\infty}^{\infty} |g(y)| f_Y(y)\, dy \\
&= \int_{\mathcal{A}_k} |g(y)| f_Y(y)\, dy + \int_{\mathcal{A}_k'} |g(y)| f_Y(y)\, dy \\
&\geq \int_{\mathcal{A}_k} |g(y)| f_Y(y)\, dy \\
&\geq \int_{\mathcal{A}_k} k f_Y(y)\, dy \\
&= k \int_{\mathcal{A}_k} f_Y(y)\, dy \\
&= kP(|g(Y)| \geq k)
\end{aligned}$$

Markov's inequality (cont.)

Thus

$$\mathbb{E}[|g(Y)|] \geq kP(|g(Y)| \geq k)$$

or equivalently

$$P(|g(Y)| \geq k) \leq \frac{\mathbb{E}[|g(Y)|]}{k}.$$

Markov's inequality (cont.)

Thus

$$1 - P(|g(Y)| \geq k) \geq 1 - \frac{\mathbb{E}[|g(Y)|]}{k}$$

or equivalently

$$P(|g(Y)| < k) \geq 1 - \frac{\mathbb{E}[|g(Y)|]}{k}.$$

Note that the result is identical for a discrete random variable.

The central limit theorem

Suppose that $Y_1, Y_2, \ldots, Y_n$, are independent, have the same distribution, with finite expectation μ and variance σ^2.

For each n, let

$$\overline{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

be the sample mean random variable.

The central limit theorem (cont.)

Then the random variable

$$U_n = \frac{\sqrt{n}(\overline{Y}_n - \mu)}{\sigma}$$

converges in distribution to a *standard normal distribution*.

That is,

$$\lim_{n \longrightarrow \infty} F_{U_n}(u) = \lim_{n \longrightarrow \infty} P(U_n \leq u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

This result is the *Central Limit Theorem* (CLT).

The central limit theorem (cont.)

Note that we could write

$$U_n = \frac{(\overline{Y}_n - \mu)}{\sigma/\sqrt{n}}$$

or, multiplying numerator and denominator by n

$$U_n = \frac{\left(\sum_{i=1}^{n} Y_i - n\mu\right)}{\sqrt{n}\sigma}$$

The central limit theorem (cont.)

The result says that, provided the conditions on expectation and variance are met, when n is large, we can approximate the distribution of $\overline{Y}_n$, or of

$$S_n = \sum_{i=1}^{n} Y_i$$

using a Normal distribution, irrespective of the actual distribution of $Y_1, \ldots, Y_n$.

The central limit theorem (cont.)

That is,

- $\overline{Y}_n$ is approximately distributed as

$$Normal(\mu, \sigma^2/n)$$

- S_n is approximately distributed as

$$Normal(n\mu, n\sigma^2)$$

The central limit theorem (cont.)

Note

If $Y_1, \ldots, Y_n$ are independent and identically distributed Normal random variables with expectation μ and variance σ^2, then

- $\overline{Y}_n$ is exactly distributed as

$$Normal(\mu, \sigma^2/n)$$

- S_n is exactly distributed as

$$Normal(n\mu, n\sigma^2)$$

The central limit theorem (cont.)

The result is straightforward to prove using mgfs. For $i = 1, 2, \ldots, n$, let

$$Z_i = \frac{Y_i - \mu}{\sigma}.$$

Then from properties of expectations and variances, we have that

$$\mathbb{E}[Z_i] = 0 \qquad \mathbb{V}[Z_i] = 1.$$

and also

$$U_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i$$

The central limit theorem (cont.)

If we assume that $m(t)$ is the mgf of each Y_i, and $m_Z(t)$ is the mgf for each Z_i, then by previous mgf results

$$m_{U_n}(t) = \left\{ m_Z\left(\frac{t}{\sqrt{n}}\right) \right\}^n$$

Now, using a Taylor expansion of $m_Z(t)$ at $t = 0$

$$m_Z(t) = m_Z(0) + t\, m_Z^{(1)}(0) + \frac{t^2}{2}\, m_Z^{(2)}(t_0)$$

where $0 < t_0 < t$ yields the remainder term.

The central limit theorem (cont.)

Now as $\mathbb{E}[Z_i] = 0$, and $\mathbb{V}[Z_i] = 1$ we can deduce that

$$m_Z(0) = 1 \qquad m_Z^{(1)}(0) = \mathbb{E}[Z_i] = 0$$

and hence

$$m_Z(t) = 1 + \frac{t^2}{2}\, m_Z^{(2)}(t_0).$$

The central limit theorem (cont.)

Therefore

$$m_{U_n}(t) = \left\{1 + \frac{t^2}{2n}\, m_Z^{(2)}(t_0)\right\}^n \qquad 0 < t_0 < \frac{t}{\sqrt{n}}$$

Now, as $n \longrightarrow \infty$

$$\frac{t}{\sqrt{n}} \longrightarrow 0$$

so $t_0 \longrightarrow 0$.

The central limit theorem (cont.)

Hence

$$\lim_{n \longrightarrow \infty} m_{U_n}(t) = \lim_{n \longrightarrow \infty} \left\{ 1 + \frac{t^2}{2n} m_Z^{(2)}(t_0) \right\}^n$$

$$= \lim_{n \longrightarrow \infty} \left\{ 1 + \frac{t^2}{2n} m_Z^{(2)}(0) \right\}^n$$

The central limit theorem (cont.)

But $m_Z^{(2)}(0) = \mathbb{E}[Z_i^2] = 1$, so therefore

$$\lim_{n \longrightarrow \infty} m_{U_n}(t) = \lim_{n \longrightarrow \infty} \left\{ 1 + \frac{t^2}{2n} \right\}^n = e^{t^2/2}$$

by definition of the exponential function. Hence the limiting mgf is that of a standard Normal distribution, and so

$$U_n \xrightarrow{d} Z$$

where Z has a standard Normal distribution.

Discrete bivariate distributions

A *discrete bivariate probability mass function* is a function of two arguments

$$p(y_1, y_2)$$

that distributes probability across the possible values of the vector (Y_1, Y_2) so that

$$p(y_1, y_2) = P((Y_1 = y_1) \cap (Y_2 = y_2)) \equiv P(Y_1 = y_1, Y_2 = y_2)$$

for $-\infty < y_1 < \infty$ and $-\infty < y_2 < \infty$.

Discrete bivariate distributions (cont.)

We can think of $p(y_1, y_2)$ as specifying the values in a probability table.

		Y_2		
		1	2	3
Y_1	1	$p(1,1)$	$p(1,2)$	$p(1,3)$
	2	$p(2,1)$	$p(2,2)$	$p(2,3)$
	3	$p(3,1)$	$p(3,2)$	$p(3,3)$
	4	$p(4,1)$	$p(4,2)$	$p(4,3)$
	5	$p(5,1)$	$p(5,2)$	$p(5,3)$

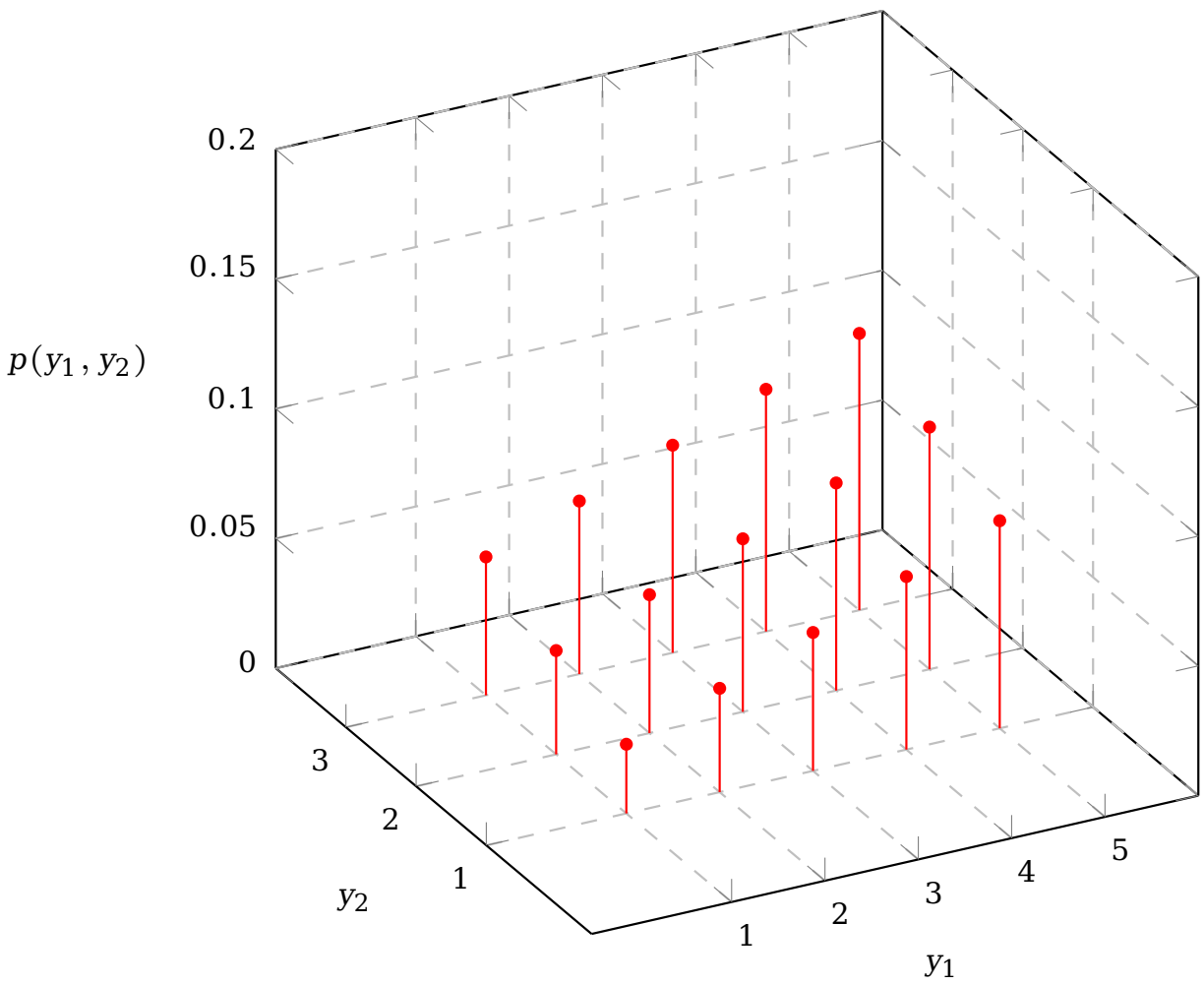
Discrete bivariate distributions (cont.)

For $1 \leq y_1 \leq 5, 1 \leq y_2 \leq 3$

$$p(y_1, y_2) = \frac{(y_1 + y_2)}{75}$$

		Y_2		
		1	2	3
Y_1	1	2/75	3/75	4/75
	2	3/75	4/75	5/75
	3	4/75	5/75	6/75
	4	5/75	6/75	7/75
	5	6/75	7/75	8/75

Discrete bivariate distributions (cont.)



Discrete bivariate distributions (cont.)

The function $p(y_1, y_2)$ is the *joint probability mass function*: it has two basic properties

- "specifies probabilities"

$$0 \leq p(y_1, y_2) \leq 1 \quad \text{for all } y_1, y_2$$

- "sums to one"

$$\sum_{y_1=-\infty}^{\infty} \sum_{y_2=-\infty}^{\infty} p(y_1, y_2) = 1.$$

although $p(y_1, y_2)$ may be zero for some arguments.

Discrete bivariate distributions (cont.)

In the above example,

$$\sum_{y_1=1}^{5}\sum_{y_2=1}^{3}\frac{(y_1+y_2)}{75} = \frac{1}{75}\sum_{y_1=1}^{5}\sum_{y_2=1}^{3}(y_1+y_2)$$

$$= \frac{1}{75}\left[3\sum_{y_1=1}^{5}y_1 + 5\sum_{y_2=1}^{3}y_2\right]$$

$$= \frac{1}{75}\left[3\frac{5\times 6}{2} + 5\frac{3\times 4}{2}\right]$$

$$= \frac{1}{75}\left[45+30\right]$$

$$= 1.$$

Discrete bivariate distributions (cont.)

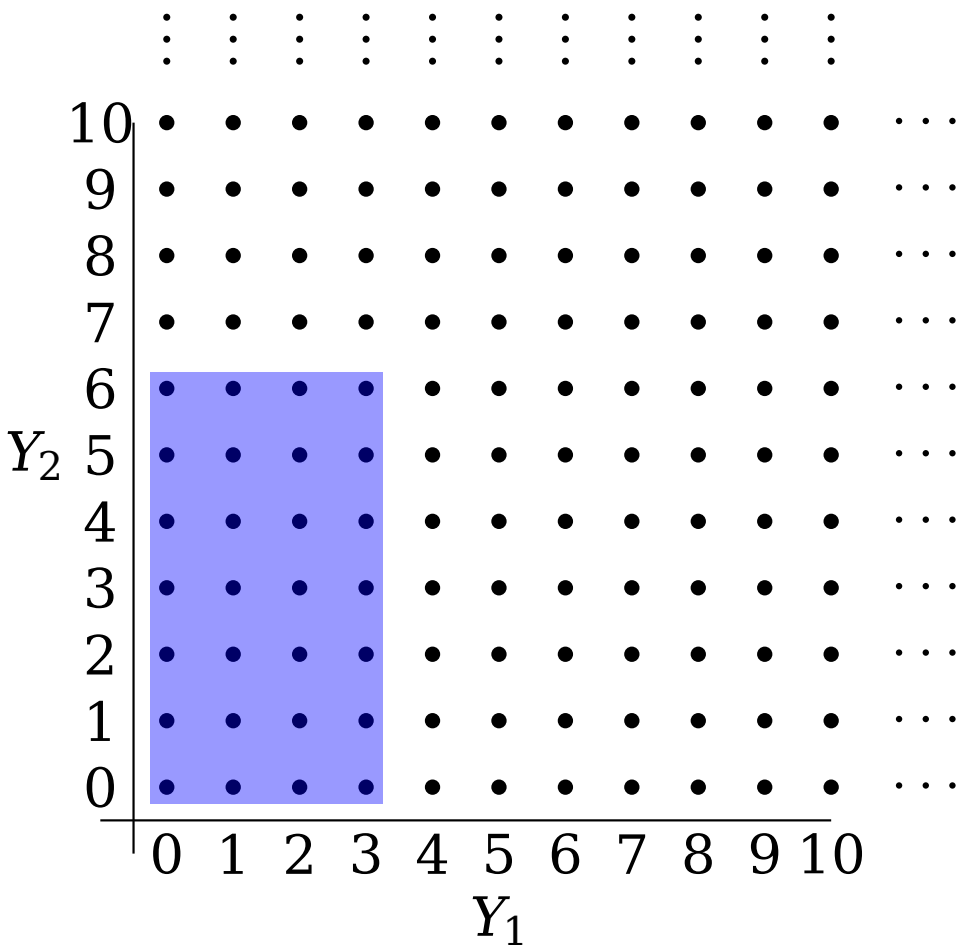
We define the *joint cumulative distribution function* $F(y_1, y_2)$ by

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2) = \sum_{t_1=-\infty}^{y_1} \sum_{t_2=-\infty}^{y_2} p(t_1, t_2)$$

that is, by accumulating the probabilities in the joint pmf over a range of values up to and including (y_1, y_2)

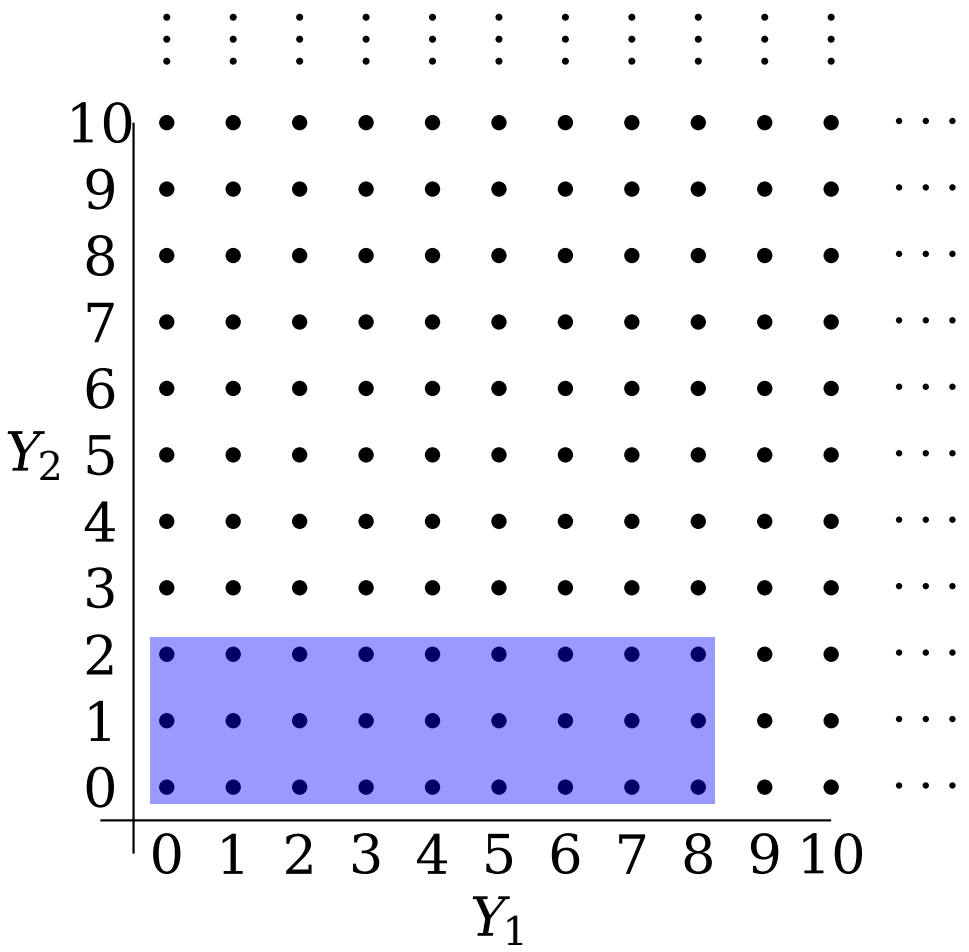
Discrete bivariate distributions (cont.)

$$F(3,6) = P(Y_1 \leq 3, Y_2 \leq 6) = \sum_{t_1=0}^{3} \sum_{t_2=0}^{6} p(t_1, t_2)$$



Discrete bivariate distributions (cont.)

$$F(8,2) = P(Y_1 \leq 8, Y_2 \leq 2) = \sum_{t_1=0}^{8} \sum_{t_2=0}^{2} p(t_1, t_2)$$



Discrete bivariate distributions (cont.)

Example

A bag contains ten balls:

- five red;
- three yellow;
- two white;

Four balls are selected, with all such selections being equally likely.

Discrete bivariate distributions (cont.)

Example

Let

- Y_1 denote the number of red balls selected;
- Y_2 denote the number of yellow balls selected.

Then using combinatorial arguments, we see that the joint pmf of Y_1 and Y_2 is given by

$$p_{Y_1,Y_2}(y_1,y_2) = \frac{\binom{5}{y_1}\binom{3}{y_2}\binom{2}{4-y_1-y_2}}{\binom{10}{4}}$$

for (y_1, y_2) such that the combinatorial terms are defined, and zero when the terms are not.

Discrete bivariate distributions (cont.)

Example

We need (y_1, y_2) simultaneously to satisfy

$$0 \leq y_1 \leq 5 \qquad 0 \leq y_2 \leq 3 \qquad 0 \leq 4 - y_1 - y_2 \leq 2$$

in order to have a non-zero probability.

Total number of selections: $\binom{10}{4} = 210.$

Discrete bivariate distributions (cont.)

Example

		Y_2			
		0	1	2	3
	0	0	0		
	1	0			
	2				0
Y_1	3			0	0
	4		0	0	0
	5	0	0	0	0

Red	Yellow	White
0	2	2
0	3	1
1	1	2
1	2	1
1	3	0
2	0	2
2	1	1
2	2	0
3	0	1
3	1	0
4	0	0

Discrete bivariate distributions (cont.)

Example

Red	Yellow	White	Count
0	2	2	3
0	3	1	2
1	1	2	15
1	2	1	30
1	3	0	5
2	0	2	10
2	1	1	60
2	2	0	30
3	0	1	20
3	1	0	30
4	0	0	5

Discrete bivariate distributions (cont.)

Example

		Y_2			
		0	1	2	3
Y_1	0	0.0000	0.0000	0.0143	0.0095
	1	0.0000	0.0714	0.1429	0.0238
	2	0.0476	0.2857	0.1429	0.0000
	3	0.0952	0.1429	0.0000	0.0000
	4	0.0238	0.0000	0.0000	0.0000
	5	0.0000	0.0000	0.0000	0.0000

Discrete bivariate distributions (cont.)

For more than two variables, we can define a joint pmf in the same way:

$$p(y_1, y_2, \ldots, y_n) = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$$

to represent an n-way table.

Discrete bivariate distributions (cont.)

Example (Previous example)

Four balls selected from 10.

- Y_1 denote the number of red balls selected;
- Y_2 denote the number of yellow balls selected.

The joint pmf of Y_1 and Y_2 is given by

$$p_{Y_1,Y_2}(y_1,y_2) = \frac{\binom{5}{y_1}\binom{3}{y_2}\binom{2}{4-y_1-y_2}}{\binom{10}{4}}$$

for (y_1, y_2) such that the combinatorial terms are defined, and zero when the terms are not.

Discrete bivariate distributions (cont.)

Example (Previous example)

We can compute the marginal pmf for Y_1 by summing probabilities in the joint probability table.

$$p_{Y_1}(0) = \frac{\binom{5}{0}\binom{3}{2}\binom{2}{2} + \binom{5}{0}\binom{3}{3}\binom{2}{1}}{\binom{10}{4}} = \frac{3+2}{210} = \frac{5}{210}$$

Discrete bivariate distributions (cont.)

Example (Previous example)

$$p_{Y_1}(1) = \frac{\binom{5}{1}\binom{3}{1}\binom{2}{2} + \binom{5}{1}\binom{3}{2}\binom{2}{1} + \binom{5}{1}\binom{3}{3}\binom{2}{0}}{\binom{10}{4}}$$

$$= \frac{15 + 30 + 5}{210} = \frac{50}{210}$$

Discrete bivariate distributions (cont.)

Example (Previous example)

$$p_{Y_1}(2) = \frac{\binom{5}{2}\binom{3}{0}\binom{2}{2} + \binom{5}{2}\binom{3}{1}\binom{2}{1} + \binom{5}{2}\binom{3}{0}\binom{2}{2}}{\binom{10}{4}}$$

$$= \frac{10 + 60 + 30}{210} = \frac{100}{210}$$

Discrete bivariate distributions (cont.)

Example (Previous example)

$$p_{Y_1}(3) = \frac{\binom{5}{3}\binom{3}{0}\binom{2}{1} + \binom{5}{3}\binom{3}{1}\binom{2}{0}}{\binom{10}{4}}$$

$$= \frac{20 + 30}{210} = \frac{50}{210}$$

Discrete bivariate distributions (cont.)

Example (Previous example)

$$p_{Y_1}(4) = \frac{\binom{5}{4}\binom{3}{0}\binom{2}{0}}{\binom{10}{4}} = \frac{5}{210}$$

Discrete bivariate distributions (cont.)

Note

In this example we can compute $p_{Y_1}(.)$ directly using the hypergeometric formula

$$p_{Y_1}(y_1) = \frac{\binom{5}{y_1}\binom{5}{4-y_1}}{\binom{10}{4}}$$

for $0 \leq y_1 \leq 5$ and $0 \leq 4 - y_1 \leq 5$.

Discrete bivariate distributions (cont.)

Note

y_1	Numerator
0	$\binom{5}{0}\binom{5}{4} = 1 \times 5 = 5$
1	$\binom{5}{1}\binom{5}{2} = 5 \times 10 = 50$
2	$\binom{5}{2}\binom{5}{3} = 10 \times 10 = 100$
3	$\binom{5}{3}\binom{5}{1} = 10 \times 5 = 50$
4	$\binom{5}{4}\binom{5}{0} = 5 \times 1 = 5$

Continuous example

Example (Distribution on the unit square)

Suppose Y_1 and Y_2 are continuous with joint pdf

$$f_{Y_1,Y_2}(y_1, y_2) = c(y_1 + y_2) \qquad 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1$$

with $f_{Y_1,Y_2}(y_1, y_2) = 0$ otherwise.

Continuous example (cont.)

Example (Distribution on the unit square)

Then for $0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1$,

$$
\begin{aligned}
F_{Y_1,Y_2}(y_1, y_2) &= \int_{-\infty}^{y_1} \left\{ \int_{-\infty}^{y_2} f_{Y_1,Y_2}(t_1, t_2)\, dt_2 \right\} dt_1 \\
&= \int_{0}^{y_1} \left\{ \int_{0}^{y_2} c(t_1 + t_2)\, dt_2 \right\} dt_1 \\
&= \int_{0}^{y_1} \left[c\left(t_1 t_2 + \frac{1}{2} t_2^2 \right) \right]_0^{y_2} dt_1 \\
&= \int_{0}^{y_1} c\left(t_1 y_2 + \frac{1}{2} y_2^2 \right) dt_1
\end{aligned}
$$

Continuous example (cont.)

Example (Distribution on the unit square)

$$= \left[c \left(\frac{1}{2} t_1^2 y_2 + \frac{1}{2} y_2^2 t_1 \right) \right]_0^{y_1}$$

$$= \frac{c}{2} \left(y_1^2 y_2 + y_1 y_2^2 \right)$$

We require that $F_{Y_1,Y_2}(1,1) = 1$, so we must have $c = 1$.

Continuous example (cont.)

Example (Distribution on the unit square)

That is

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} (y_1+y_2) & 0 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(no probability outside of the unit square).

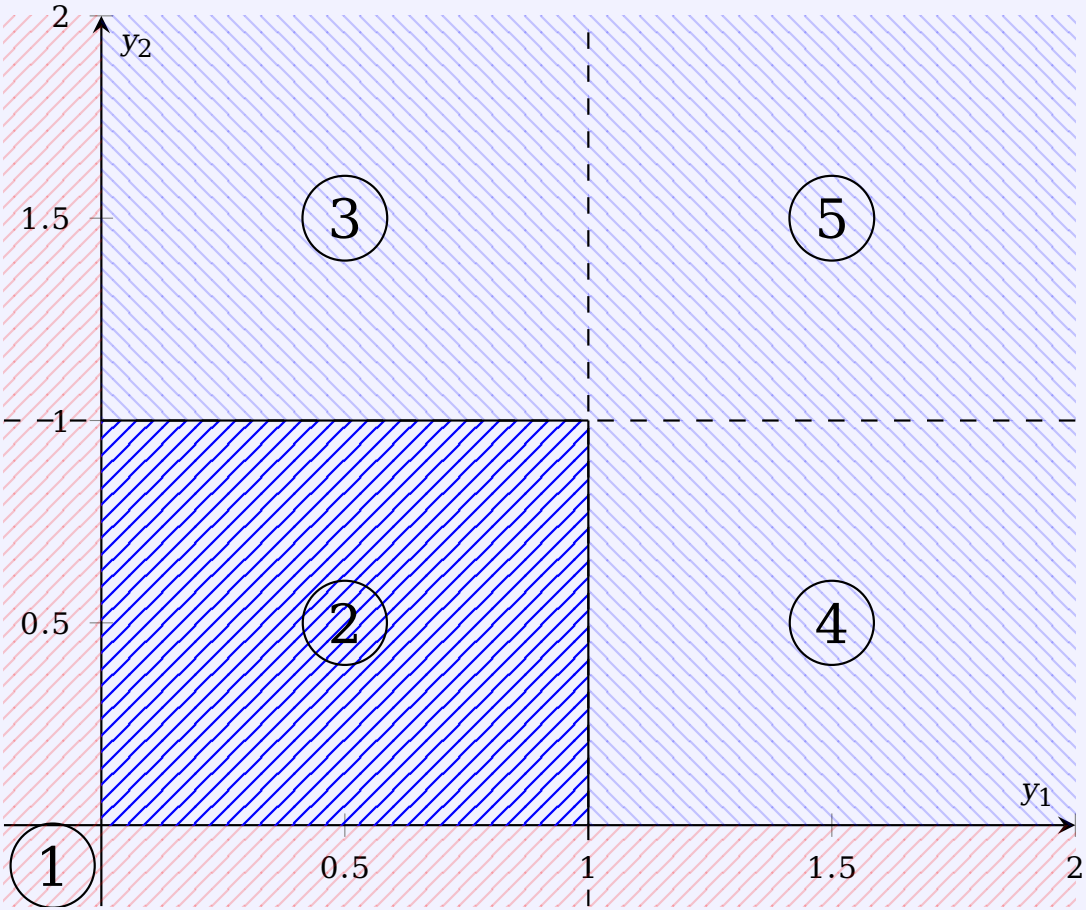
Continuous example (cont.)

Example (Distribution on the unit square)

$$F_{Y_1,Y_2}(y_1,y_2) = \begin{cases} 0 & y_1 < 0 \text{ or } y_2 < 0 \quad (1) \\ \left(y_1^2 y_2 + y_1 y_2^2\right)/2 & 0 \le y_1, y_2 \le 1 \quad (2) \\ \left(y_1^2 + y_1\right)/2 & 0 \le y_1 \le 1, y_2 > 1 \quad (3) \\ \left(y_2 + y_2^2\right)/2 & 0 \le y_2 \le 1, y_1 > 1 \quad (4) \\ 1 & y_1 > 1 \text{ and } y_2 > 1 \quad (5) \end{cases}$$

Continuous example (cont.)

Note: regions for $F_{Y_1,Y_2}(y_1, y_2)$



Continuous example (cont.)

Note

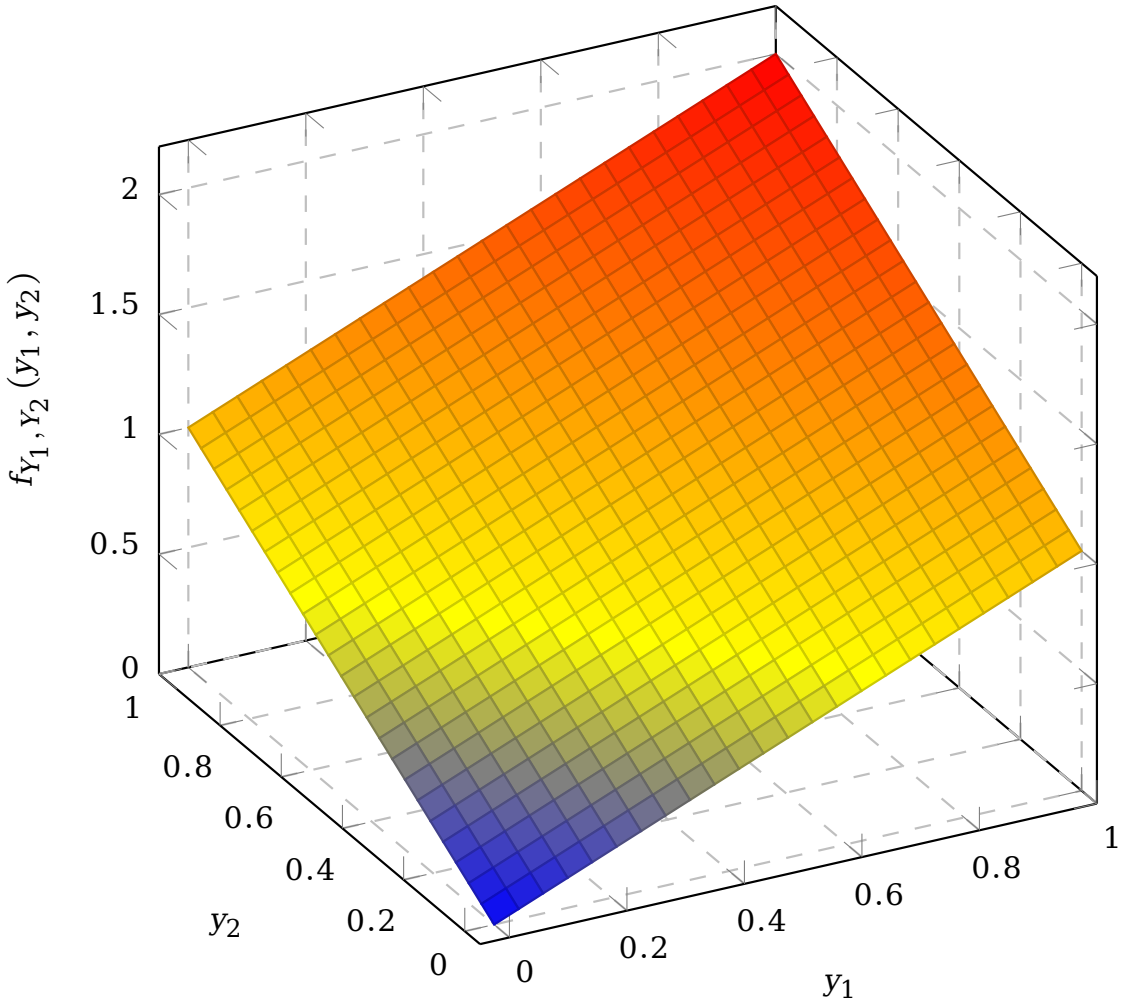
To compute $F_{Y_1,Y_2}(y_1, y_2)$ we always integrate the joint pdf

below and to the left

of (y_1, y_2).

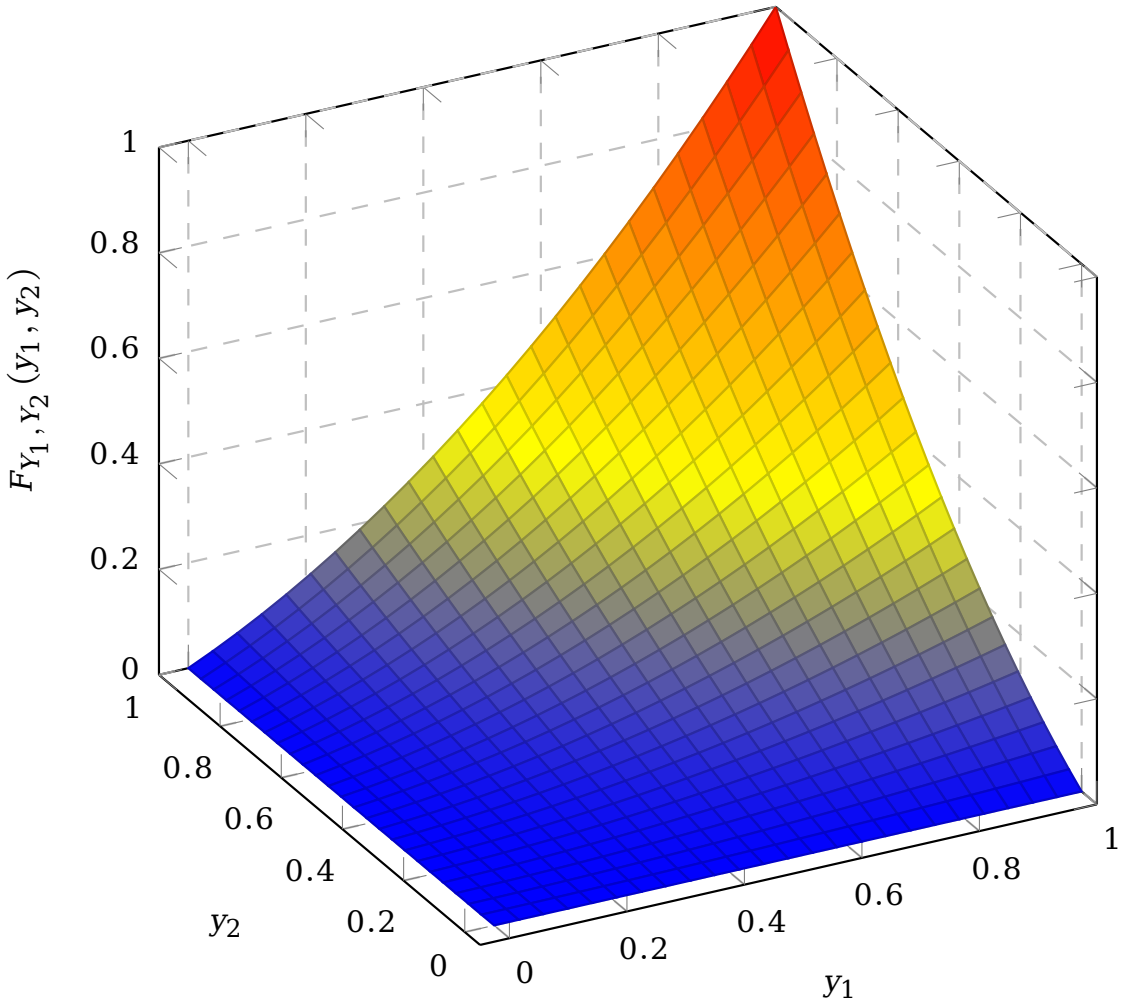
Continuous example (cont.)

$$f_{Y_1,Y_2}(y_1, y_2) = (y_1 + y_2)$$



Continuous example (cont.)

$$F_{Y_1,Y_2}(y_1, y_2) = (y_1^2 y_2 + y_1 y_2^2)/2$$



Continuous example (cont.)

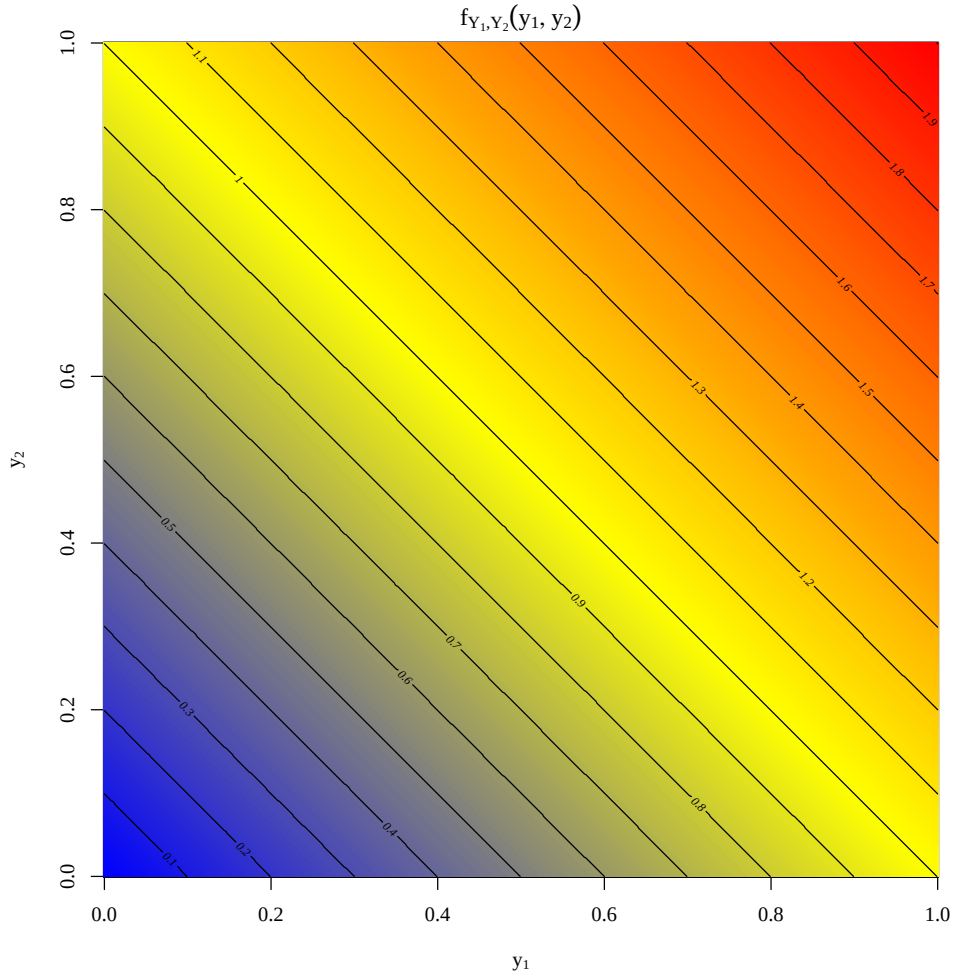


Figure: $f_{Y_1,Y_2}(y_1, y_2)$: image and contour plot

Continuous example (cont.)

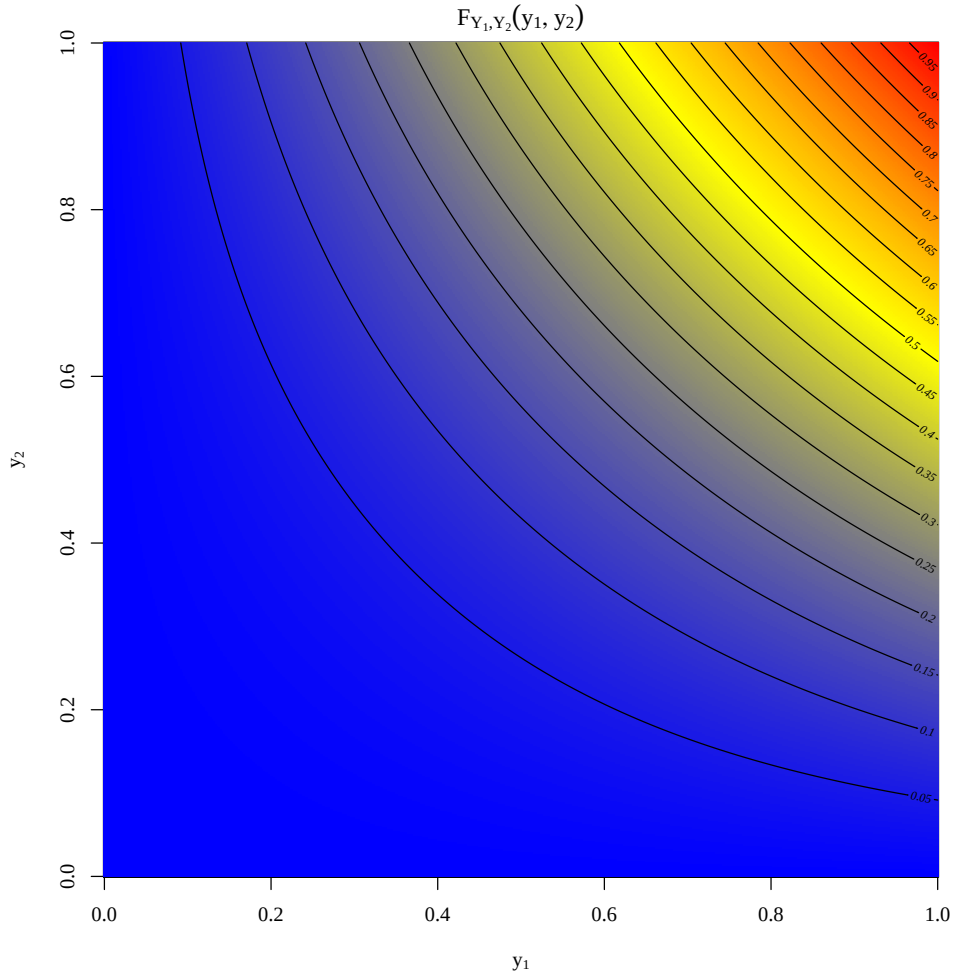


Figure: $F_{Y_1,Y_2}(y_1, y_2)$: image and contour plot

Continuous example (cont.)

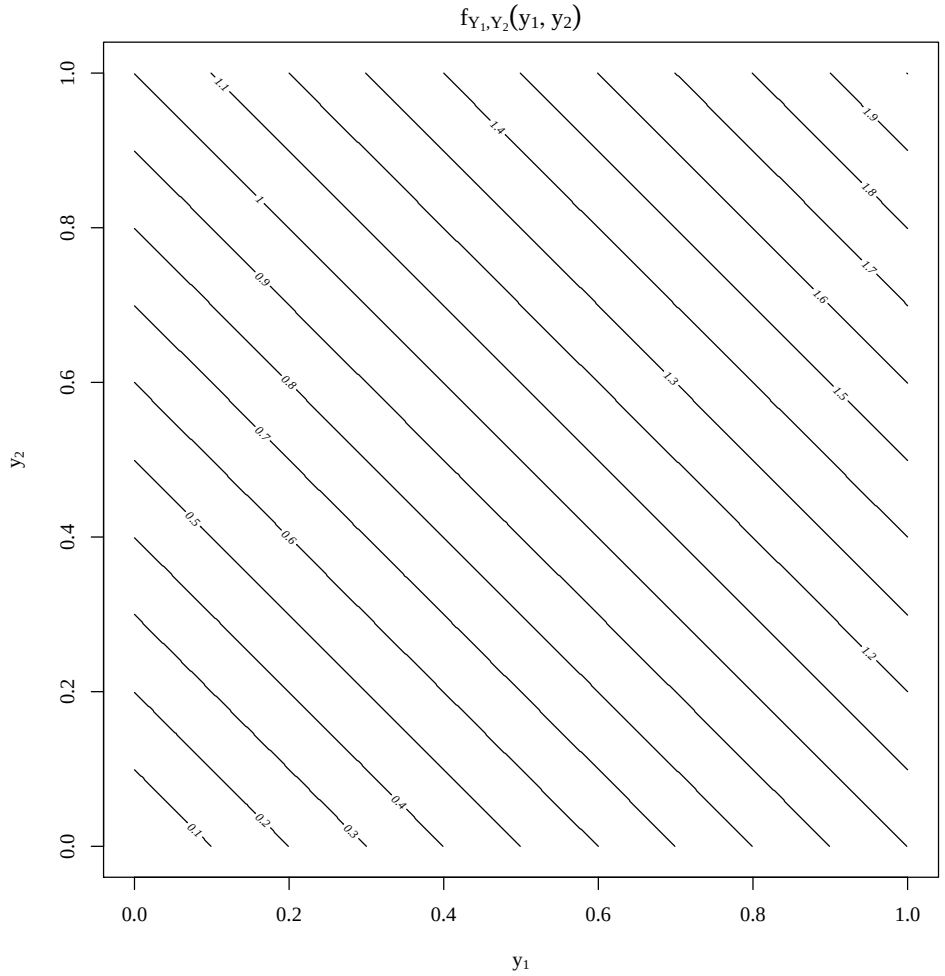


Figure: $f_{Y_1,Y_2}(y_1, y_2)$: contour plot

Continuous example (cont.)

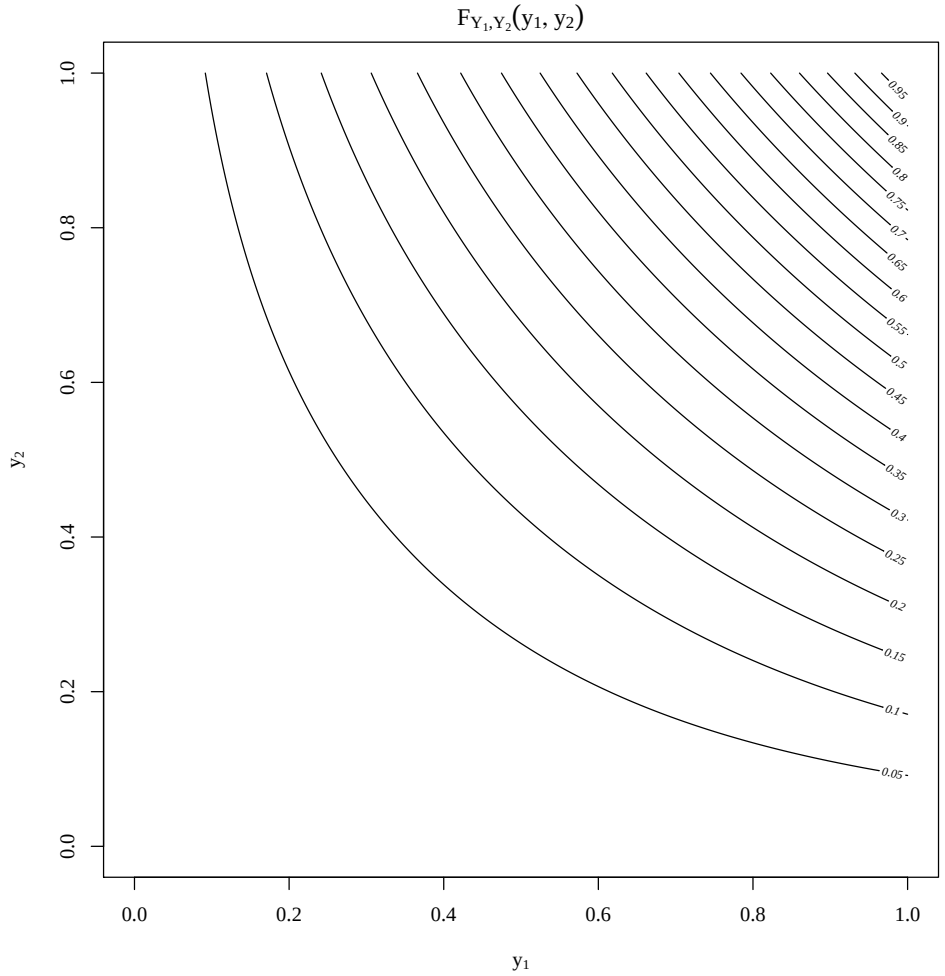


Figure: $F_{Y_1,Y_2}(y_1, y_2)$: contour plot

Partial differentiation

Example

For $0 \leq y_1, y_2 \leq 1$,

$$F_{Y_1,Y_2}(y_1, y_2) = \frac{1}{2}\left(y_1^2 y_2 + y_1 y_2^2\right)$$

- **Step 1:**

$$\frac{\partial F_{Y_1,Y_2}(y_1, y_2)}{\partial y_2} = \frac{1}{2}\left(y_1^2 + 2y_1 y_2\right)$$

- **Step 2:**

$$\frac{\partial\left\{\frac{1}{2}\left(y_1^2 + 2y_1 y_2\right)\right\}}{\partial y_2} = \frac{1}{2}\left(2y_1 + 2y_2\right) = \left(y_1 + y_2\right)$$

The marginal density function

Once the joint pdf $f_{Y_1,Y_2}(y_1, y_2)$ is specified, we define the *marginal probability density functions* (or marginal pdfs) in a similar way as for the discrete case.

We have for Y_1 the marginal pdf $f_{Y_1}(y_1)$ for each fixed y_1 by

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2)\, dy_2$$

and for Y_2 the marginal pdf $f_{Y_2}(y_2)$ for each fixed y_2 by

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2)\, dy_1$$

The marginal density function (cont.)

Example (Distribution on the unit square)

For

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} (y_1+y_2) & 0 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

The marginal density function (cont.)

Example (Distribution on the unit square)

We have for $0 \leq y_1 \leq 1$

$$\begin{aligned} f_{Y_1}(y_1) &= \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2)\, dy_2 \\ &= \int_0^1 (y_1 + y_2)\, dy_2 \\ &= \left[y_1 y_2 + \frac{y_2^2}{2} \right]_0^1 \\ &= y_1 + \frac{1}{2}. \end{aligned}$$

with $f_{Y_1}(y_1) = 0$ otherwise.

The marginal density function (cont.)

Example (Distribution on the unit square)

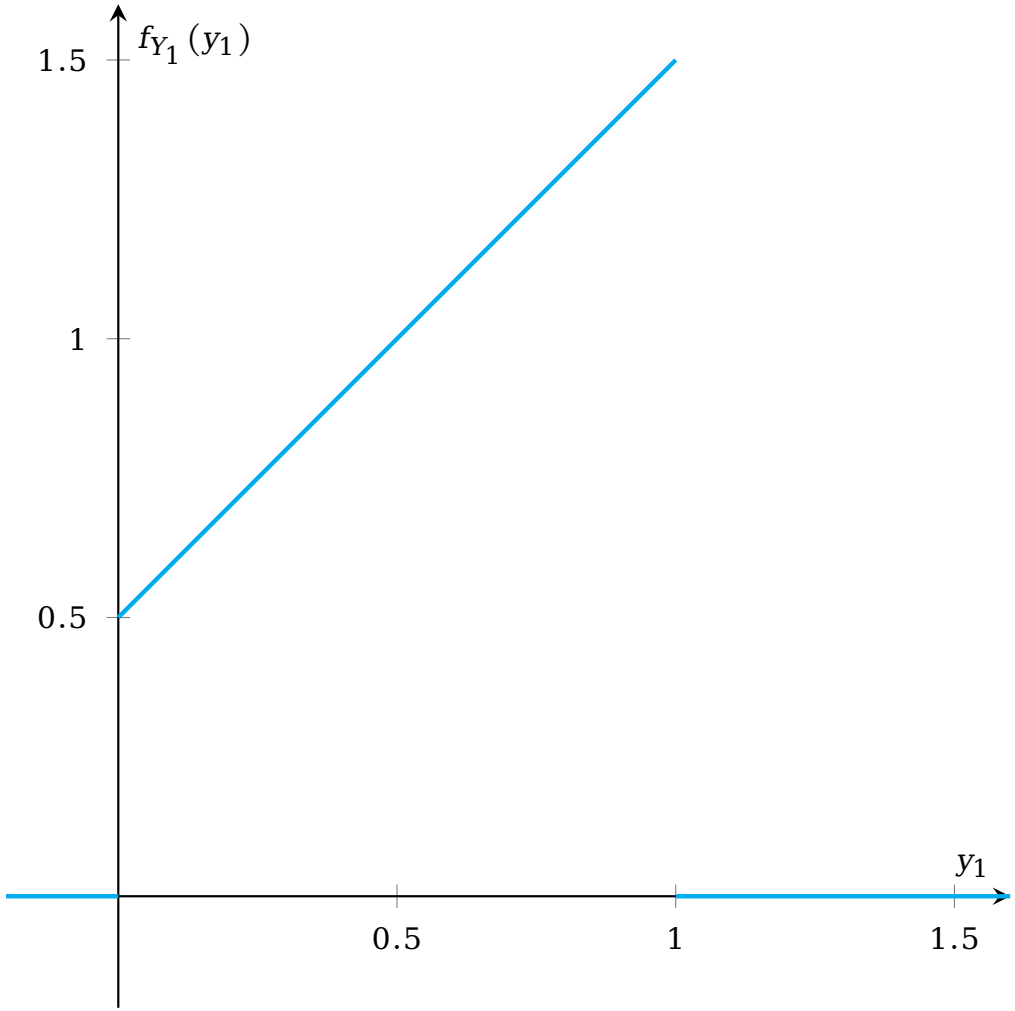
Also for $0 \leq y_1 \leq 1$

$$\begin{aligned} F_{Y_1}(y_1) &= \int_{-\infty}^{y_1} f_{Y_1}(t_1)\, dt_1 \\ &= \int_0^{y_1} \left(t_1 + \frac{1}{2} \right) \, dt_2 \\ &= \left[\frac{t_1^2}{2} + \frac{t_1}{2} \right]_0^{y_1} \\ &= \frac{y_1^2}{2} + \frac{y_1}{2}. \end{aligned}$$

with $F_{Y_1}(y_1) = 0$ for $y_1 < 0$ and $F_{Y_1}(y_1) = 1$ for $y_1 > 1$.

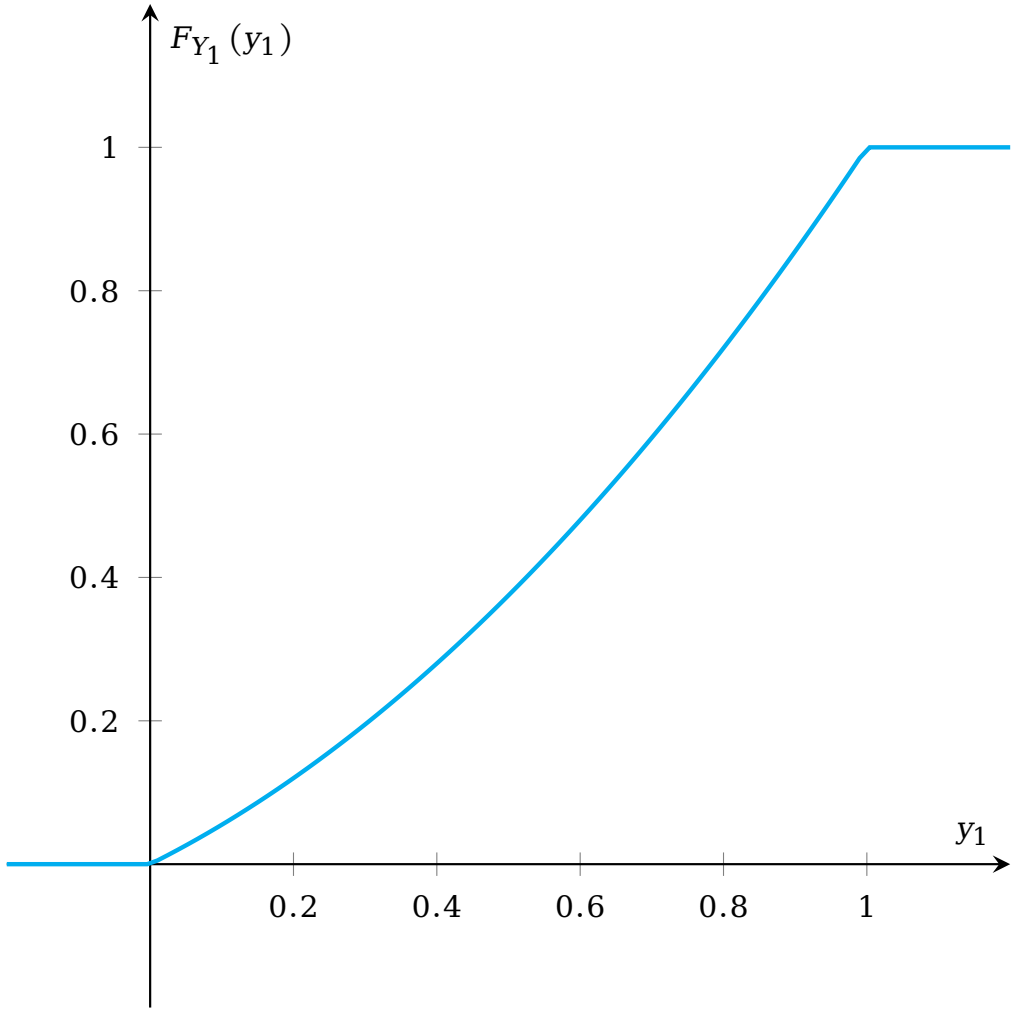
The marginal density function (cont.)

$f_{Y_1}(y_1) = y_1 + 1/2$ for $0 \leq y_1 \leq 1$.



The marginal density function (cont.)

$F_{Y_1}(y_1) = (y_1^2 + y_1)/2$ for $0 \leq y_1 \leq 1$.



The marginal density function (cont.)

Example (Distribution on the unit square)

We can perform a similar calculation and obtain

$$f_{Y_2}(y_2) = y_2 + \frac{1}{2} \qquad 0 \leq y_2 \leq 1$$

with $f_{Y_2}(y_2) = 0$ otherwise, and

$$F_{Y_2}(y_2) = \frac{(y_2^2 + y_2)}{2} \qquad 0 \leq y_2 \leq 1$$

with $F_{Y_2}(y_2) = 0$ for $y_2 < 0$ and $F_{Y_2}(y_2) = 1$ for $y_2 > 1$.

The conditional density function

Example (Distribution on the unit square)

For

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} (y_1+y_2) & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

we can consider conditioning on values of y_1 or y_2 in the interval $[0,1]$.

The conditional density function (cont.)

Example (Distribution on the unit square)

As

$$f_{Y_1}(y_1) = y_1 + \frac{1}{2} \qquad 0 \leq y_1 \leq 1.$$

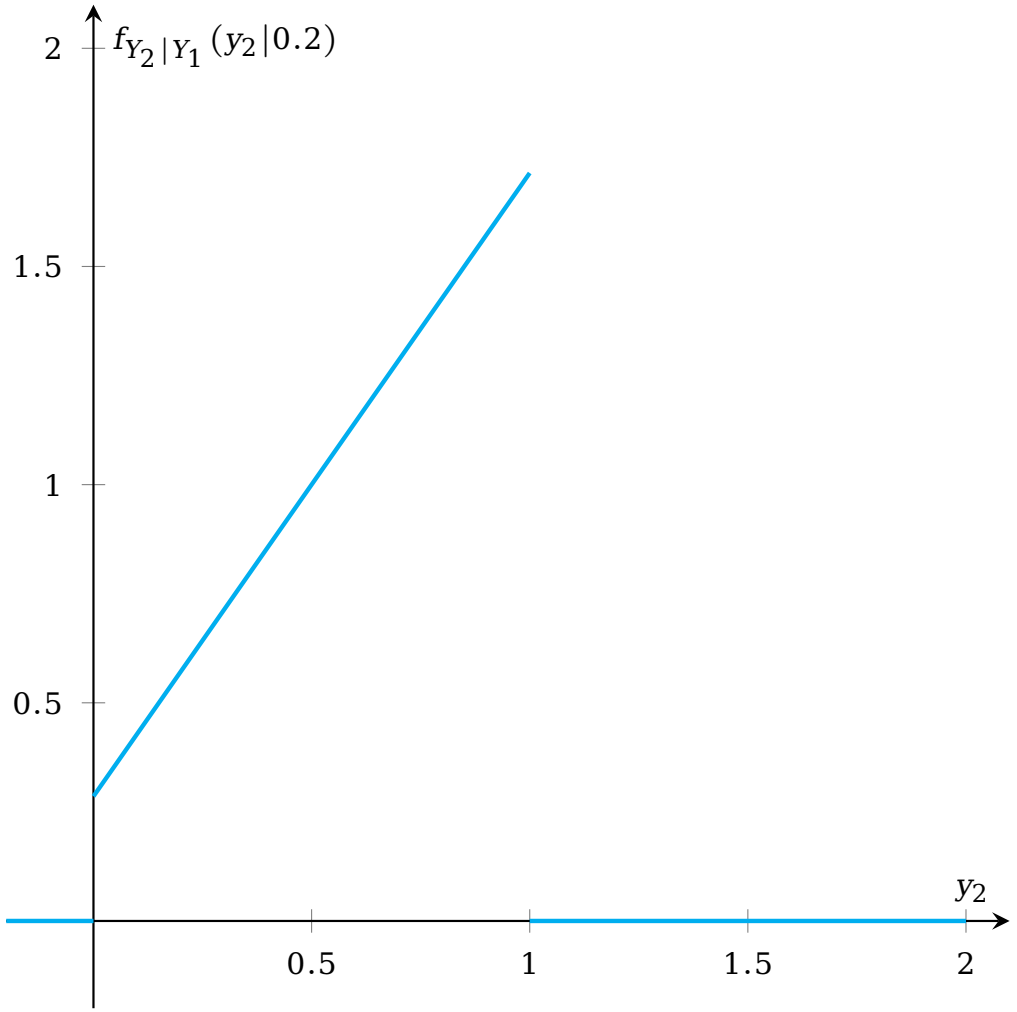
we have for each fixed y_1 in this range

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{f_{Y_1,Y_2}(y_1, y_2)}{f_{Y_1}(y_1)} = \frac{(y_1 + y_2)}{y_1 + 1/2}$$

for $0 \leq y_2 \leq 1$, with the function zero for other values of y_2.

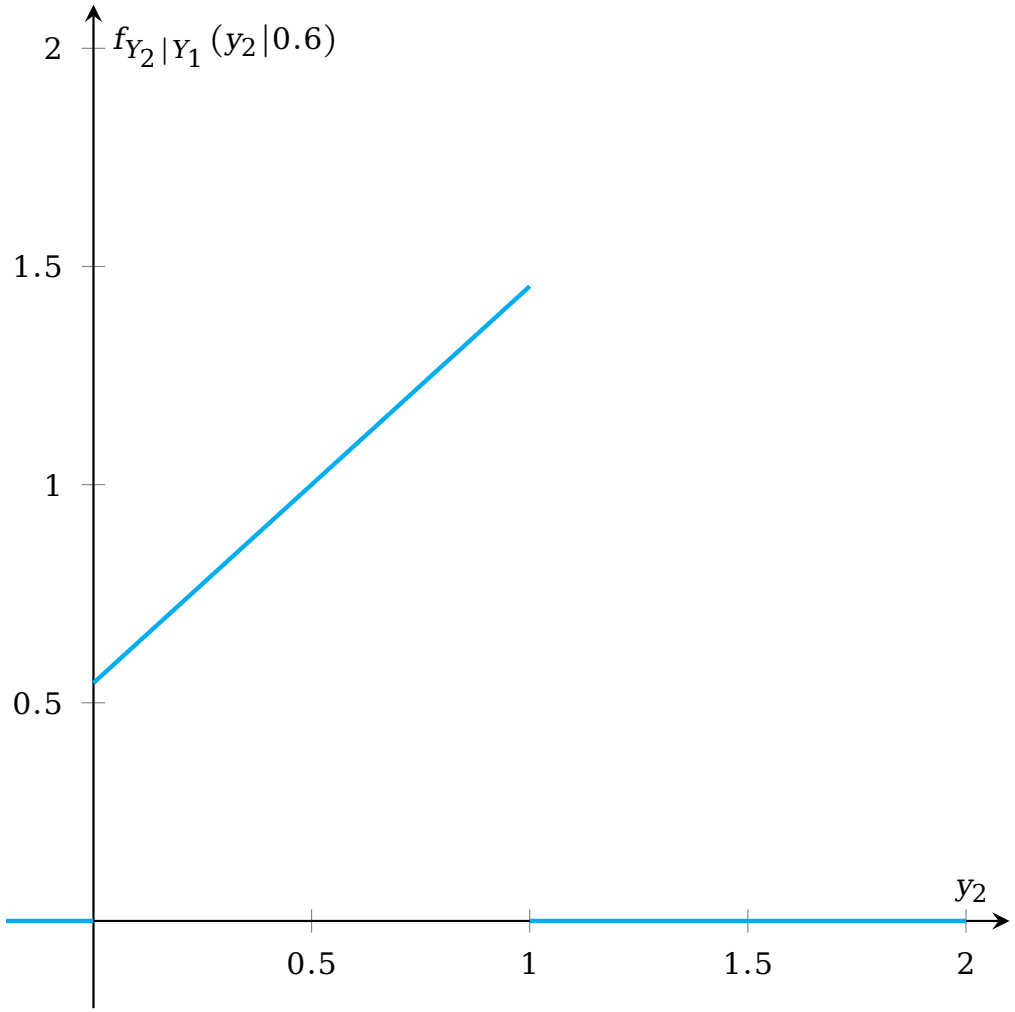
The conditional density function (cont.)

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{(y_1 + y_2)}{y_1 + 1/2} \text{ for } 0 \leq y_2 \leq 1, \text{ with } y_1 = 0.2$$



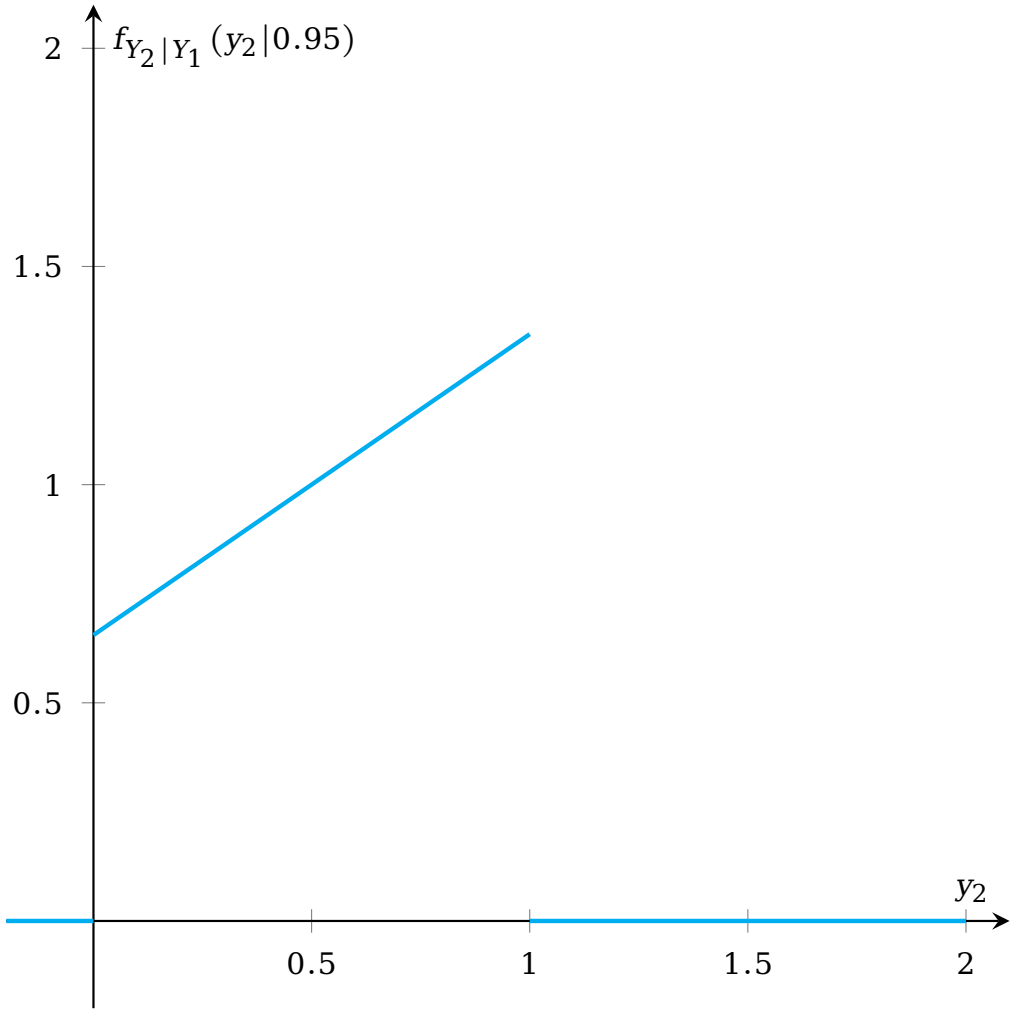
The conditional density function (cont.)

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{(y_1 + y_2)}{y_1 + 1/2} \text{ for } 0 \leq y_2 \leq 1, \text{ with } y_1 = 0.6$$



The conditional density function (cont.)

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{(y_1 + y_2)}{y_1 + 1/2} \text{ for } 0 \leq y_2 \leq 1, \text{ with } y_1 = 0.95$$



Covariance and correlation

For two random variables Y_1 and Y_2, with expectations μ_1 and μ_2 respectively, the *covariance* between Y_1 and Y_2 is written

$$\text{Cov}[Y_1, Y_2]$$

and is defined by

$$\text{Cov}[Y_1, Y_2] = \mathbb{E}[(Y_1 - \mu_1)(Y_2 - \mu_2)]$$

Covariance and correlation (cont.)

Note that

$$\begin{aligned}
\mathrm{Cov}[Y_1, Y_2] &= \mathbb{E}[(Y_1 - \mu_1)(Y_2 - \mu_2)] \\
&= \mathbb{E}[Y_1 Y_2 - \mu_1 Y_2 - Y_1 \mu_2 + \mu_1 \mu_2] \\
&= \mathbb{E}[Y_1 Y_2] - \mu_1 \mathbb{E}[Y_2] - \mu_2 \mathbb{E}[Y_1] + \mu_1 \mu_2 \\
&= \mathbb{E}[Y_1 Y_2] - \mu_1 \mu_2 - \mu_2 \mu_1 + \mu_1 \mu_2 \\
&= \mathbb{E}[Y_1 Y_2] - \mu_1 \mu_2 .
\end{aligned}$$

Covariance and correlation (cont.)

The *correlation* between Y_1 and Y_2, $\text{Corr}(Y_1, Y_2)$, is defined by

$$\text{Corr}[Y_1, Y_2] = \frac{\text{Cov}[Y_1, Y_2]}{\sqrt{\mathbb{V}[Y_1]\mathbb{V}[Y_2]}}$$

Covariance and correlation (cont.)

Both covariance and correlation are *measures of dependence*.
We have that for any pair of random variables,

$$-1 \leq \text{Corr}[Y_1, Y_2] \leq 1$$

Covariance and correlation (cont.)

If

$$\mathrm{Cov}[Y_1, Y_2] = \mathrm{Corr}[Y_1, Y_2] = 0$$

we say that Y_1 and Y_2 are *uncorrelated*.

Independent random variables

If Y_1 and Y_2 are independent then (in the continuous case – discrete case works in the same way), we have that

$$\begin{aligned}
\mathbb{E}[Y_1 Y_2] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 f_{Y_1,Y_2}(y_1, y_2)\, dy_1\, dy_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 f_{Y_1}(y_1) f_{Y_2}(y_2)\, dy_1\, dy_2 \\
&= \int_{-\infty}^{\infty} y_2 f_{Y_2}(y_2) \left\{ \int_{-\infty}^{\infty} y_1 f_{Y_1}(y_1)\, dy_1 \right\} dy_2 \\
&= \int_{-\infty}^{\infty} y_2 f_{Y_2}(y_2) \mathbb{E}[Y_1] dy_2 \\
&= \mathbb{E}[Y_1] \int_{-\infty}^{\infty} y_2 f_{Y_2}(y_2) dy_2 = \mathbb{E}[Y_1]\mathbb{E}[Y_2]
\end{aligned}$$

Independent random variables (cont.)

More generally, if $g_1(Y_1)$ and $g_2(Y_2)$ are two functions of the independent random variables, then

$$\mathbb{E}[g_1(Y_1)g_2(Y_2)] = \mathbb{E}[g_1(Y_1)]\mathbb{E}[g_2(Y_2)]$$

Independent random variables (cont.)

Therefore if two random variables Y_1 and Y_2 are independent, then

$$\begin{aligned}\mathrm{Cov}[Y_1, Y_2] &= \mathbb{E}[Y_1 Y_2] - \mu_1 \mu_2 \\ &= \mathbb{E}[Y_1]\mathbb{E}[Y_2] - \mu_1 \mu_2 \\ &= \mu_1 \mu_2 - \mu_1 \mu_2 = 0\end{aligned}$$

and Y_1 and Y_2 are uncorrelated.

Linear combinations of random variables

Suppose that $Y_1, \ldots, Y_n$ and $X_1, \ldots, X_m$ be random variables with

$$\mathbb{E}[Y_i] = \mu_i \qquad \mathbb{E}[X_j] = \xi_j.$$

Define

$$U_1 = \sum_{i=1}^{n} a_i Y_i \qquad U_2 = \sum_{j=1}^{m} b_j X_j$$

for real constants $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$.

Linear combinations of random variables (cont.)

(a) Expectations:

$$\mathbb{E}[U_1] = \sum_{i=1}^{n} a_i \mathbb{E}[Y_i] = \sum_{i=1}^{n} a_i \mu_i$$

(b) Variances:

$$\mathbb{V}[U_1] = \sum_{i=1}^{n} a_i^2 \mathbb{V}[Y_i] + 2 \sum_{j=2}^{n} \sum_{i=1}^{j-1} a_i a_j \mathrm{Cov}[Y_i, Y_j]$$

(c) Covariances:

$$\mathrm{Cov}[U_1, U_2] = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \mathrm{Cov}[Y_i, X_j]$$

Linear combinations of random variables (cont.)

For (b):

$$\mathbb{V}[U_1] = \mathbb{E}[U_1^2] - \{\mathbb{E}[U_1]\}^2.$$

But

$$U_1^2 = \left(\sum_{i=1}^{n} a_i Y_i\right)^2 = \sum_{i=1}^{n} a_i^2 Y_i^2 + 2\sum_{j=2}^{n}\sum_{i=1}^{j-1} a_i a_j Y_i Y_j$$

so therefore

$$\mathbb{E}[U_1^2] = \sum_{i=1}^{n} a_i^2 \mathbb{E}[Y_i^2] + 2\sum_{j=2}^{n}\sum_{i=1}^{j-1} a_i a_j \mathbb{E}[Y_i Y_j]$$

Linear combinations of random variables (cont.)

Similarly

$$\{\mathbb{E}[U_1]\}^2 = \left\{\sum_{i=1}^{n} a_i \mu_i\right\}^2 = \sum_{i=1}^{n} a_i^2 \mu_i^2 + 2\sum_{j=2}^{n}\sum_{i=1}^{j-1} a_i a_j \mu_i \mu_j$$

Therefore

$$\begin{aligned}\mathbb{E}[U_1^2] - \{\mathbb{E}[U_1]\}^2 = &\sum_{i=1}^{n} a_i^2(\mathbb{E}[Y_i^2] - \mu_i^2) \\ &+ 2\sum_{j=2}^{n}\sum_{i=1}^{j-1} a_i a_j(\mathbb{E}[Y_i Y_j] - \mu_i \mu_j)\end{aligned}$$

and the result follows.

Linear combinations of random variables (cont.)

For (c): consider the simple case

$$U_1 = a_1 Y_1 + a_2 Y_2 \qquad U_2 = b_1 X_1 + b_2 X_2$$

Then

$$\begin{aligned}
\mathbb{E}[U_1 U_2] &= \mathbb{E}[(a_1 Y_1 + a_2 Y_2)(b_1 X_1 + b_2 X_2)] \\
&= \mathbb{E}[a_1 b_1 Y_1 X_1 + a_1 b_2 Y_1 X_2 + a_2 b_1 Y_2 X_1 + a_2 b_2 Y_2 X_2] \\
&= \sum_{i=1}^{2} \sum_{j=1}^{2} a_i b_j \mathbb{E}[Y_i X_j]
\end{aligned}$$

Linear combinations of random variables (cont.)

Using the same logic

$$\begin{aligned}
\mathbb{E}[U_1]\mathbb{E}[U_2] &= \left(\sum_{i=1}^{2} a_i \mathbb{E}[Y_i]\right)\left(\sum_{j=1}^{2} b_j \mathbb{E}[X_j]\right) \\
&= \left(\sum_{i=1}^{2} a_i \mu_i\right)\left(\sum_{j=1}^{2} b_j \xi_j\right) \\
&= \sum_{i=1}^{2}\sum_{j=1}^{2} a_i b_j \mu_i \xi_j
\end{aligned}$$

Linear combinations of random variables (cont.)

and therefore

$$\mathbb{E}[U_1 U_2] - \mathbb{E}[U_1]\mathbb{E}[U_2] = \sum_{i=1}^{2}\sum_{j=1}^{2} a_i b_j (\mathbb{E}[Y_i X_j] - \mu_i \xi_j)$$

and the result follows.

Independent random variables

$$
\begin{aligned}
\mathbb{E}[Y_1 Y_2] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 f_{Y_1, Y_2}(y_1, y_2)\, dy_1\, dy_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 f_{Y_1}(y_1) f_{Y_2}(y_2)\, dy_1\, dy_2 \\
&= \int_{-\infty}^{\infty} y_2 f_{Y_2}(y_2) \left\{ \int_{-\infty}^{\infty} y_1 f_{Y_1}(y_1)\, dy_1 \right\} dy_2 \\
&= \int_{-\infty}^{\infty} y_2 f_{Y_2}(y_2) \mathbb{E}[Y_1] dy_2 \\
&= \mathbb{E}[Y_1] \int_{-\infty}^{\infty} y_2 f_{Y_2}(y_2) dy_2 = \mathbb{E}[Y_1]\mathbb{E}[Y_2]
\end{aligned}
$$

Sums of random variables (revisited)

Example (Sum of Binomial random variables)

If Y_1 and Y_2 are independent Binomial random variables with parameters (n_1, p) and (n_2, p), we have

$$m_{Y_1}(t) = (pe^t + q)^{n_1} \qquad m_{Y_2}(t) = (pe^t + q)^{n_2}$$

then if $Y = Y_1 + Y_2$, we have that

$$m_Y(t) = m_{Y_1}(t)m_{Y_2}(t) = (pe^t+q)^{n_1}(pe^t+q)^{n_2} = (pe^t+q)^{n_1+n_2}$$

and we recognize this as the mgf of the Binomial distribution with parameters $(n_1 + n_2, p)$.

Sums of random variables (revisited) (cont.)

Example (Sum of Poisson random variables)

If Y_1 and Y_2 are independent Poisson random variables with parameters λ_1 and λ_2, we have

$$m_{Y_1}(t) = \exp\{\lambda_1(e^t - 1)\} \qquad m_{Y_2}(t) = \exp\{\lambda_2(e^t - 1)\}$$

then if $Y = Y_1 + Y_2$, we have that

$$\begin{aligned} m_Y(t) = m_{Y_1}(t)m_{Y_2}(t) &= \exp\{\lambda_1(e^t - 1)\}\exp\{\lambda_2(e^t - 1)\} \\ &= \exp\{(\lambda_1 + \lambda_2)(e^t - 1)\} \end{aligned}$$

and we recognize this as the mgf of the Poisson distribution with parameter $\lambda_1 + \lambda_2$.

Sums of random variables (revisited) (cont.)

Example (Sum of Normal random variables)

If Y_1 and Y_2 are independent Normal random variables with parameters (μ_1, σ_1^2) and (μ_2, σ_2^2), we have

$$m_{Y_1}(t) = \exp\{\mu_1 t + \sigma_1^2 t^2/2\} \qquad m_{Y_2}(t) = \exp\{\mu_2 t + \sigma_2^2 t^2/2\}$$

then if $Y = Y_1 + Y_2$, we have that

$$\begin{aligned} m_Y(t) = m_{Y_1}(t) m_{Y_2}(t) &= \exp\{\mu_1 t + \sigma_1^2 t^2/2\} \exp\{\mu_2 t + \sigma_2^2 t^2/2\} \\ &= \exp\{(\mu_1 + \mu_2)t + (\sigma_1^2 + \sigma_2^2)t^2/2\} \end{aligned}$$

and we recognize this as the mgf of the Normal distribution with parameters $(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.