37262 Mathematical Statistics

Lecture 3

Univariate Distributions

- So far, we have only looked at distributions of a single random variable. These are **univariate distributions**.

- For a discrete random variable, we can define the distribution by its associated probability mass function.

- For a continuous random variable, we can define the distribution by its associated probability density function.

- In many situations, however, we have multiple variables defined from the same random experiment.

Multivariate Distributions

- Consider the random experiment of flipping three fair coins and recording the outcomes.

- Let X be the number of Tails observed on the first two flips.

- Let Y be the number of Tails observed on the first three flips.

- The **joint distribution** of X and Y is the probability distribution that the ordered pair (X,Y) takes a given pair of values.

- For the discrete case here, we have the joint probability mass function $P((X,Y)=(m,n))$ over all possible pairs (m,n).

Multivariate Distributions

- For this experiment, the range of X is {0, 1, 2).

- The range of Y is {0, 1, 2, 3}.

- Note, though, that not all pairs of (X,Y) are possible. For example $P((X,Y)=(1,0))=0$ since we cannot have one Tails in the first two flips but zero in the first three.

- Here, Y can only take values equal to X or one larger than it, since the third flip either adds zero Tails to the total or it adds one Tails to the total.

Multivariate Distributions

- $X \sim Bin(2,0.5)$ hence $P(X = m) = \begin{cases} 0.25 & m = 0 \\ 0.5 & m = 1 \\ 0.25 & m = 2 \\ 0 & \text{otherwise} \end{cases}$

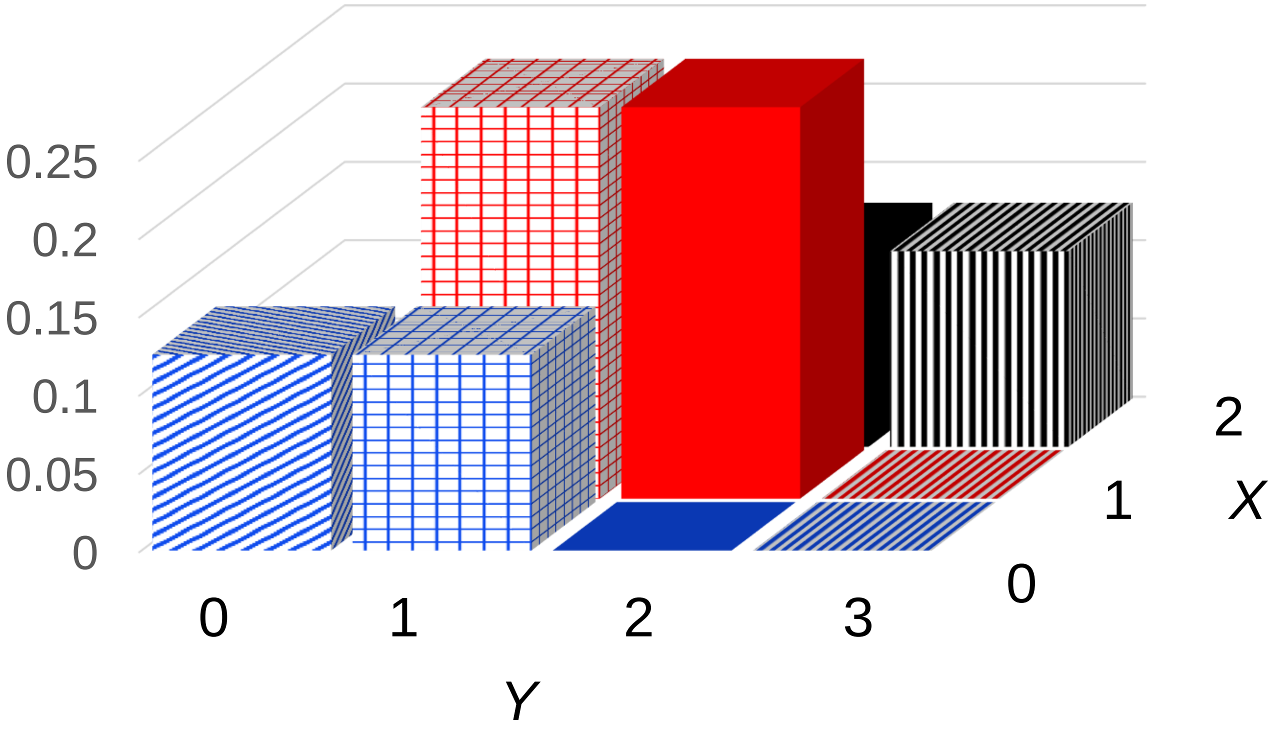
- Half of the time X and Y are the same,
  and half of the time Y is one larger, hence $P((X,Y) = (m,n)) = \begin{cases} 0.125 & (m,n) = (0,0) \\ 0.125 & (m,n) = (0,1) \\ 0.25 & (m,n) = (1,1) \\ 0.25 & (m,n) = (1,2) \\ 0.125 & (m,n) = (2,2) \\ 0.125 & (m,n) = (2,3) \\ 0 & \text{otherwise} \end{cases}$

- Note, as with univariate cases that the probability masses must be non-negative and must sum to one.

Multivariate Distributions

- We can visualise a **bivariate** distribution (i.e. a multivariate distribution with two dimensions) with three dimensional plot, similar to how we visualise univariate distributions.

$$P((X,Y)=(m,n))=\begin{cases} 0.125 & (m,n)=(0,0) \\ 0.125 & (m,n)=(0,1) \\ 0.25 & (m,n)=(1,1) \\ 0.25 & (m,n)=(1,2) \\ 0.125 & (m,n)=(2,2) \\ 0.125 & (m,n)=(2,3) \\ 0 & \text{otherwise} \end{cases}$$

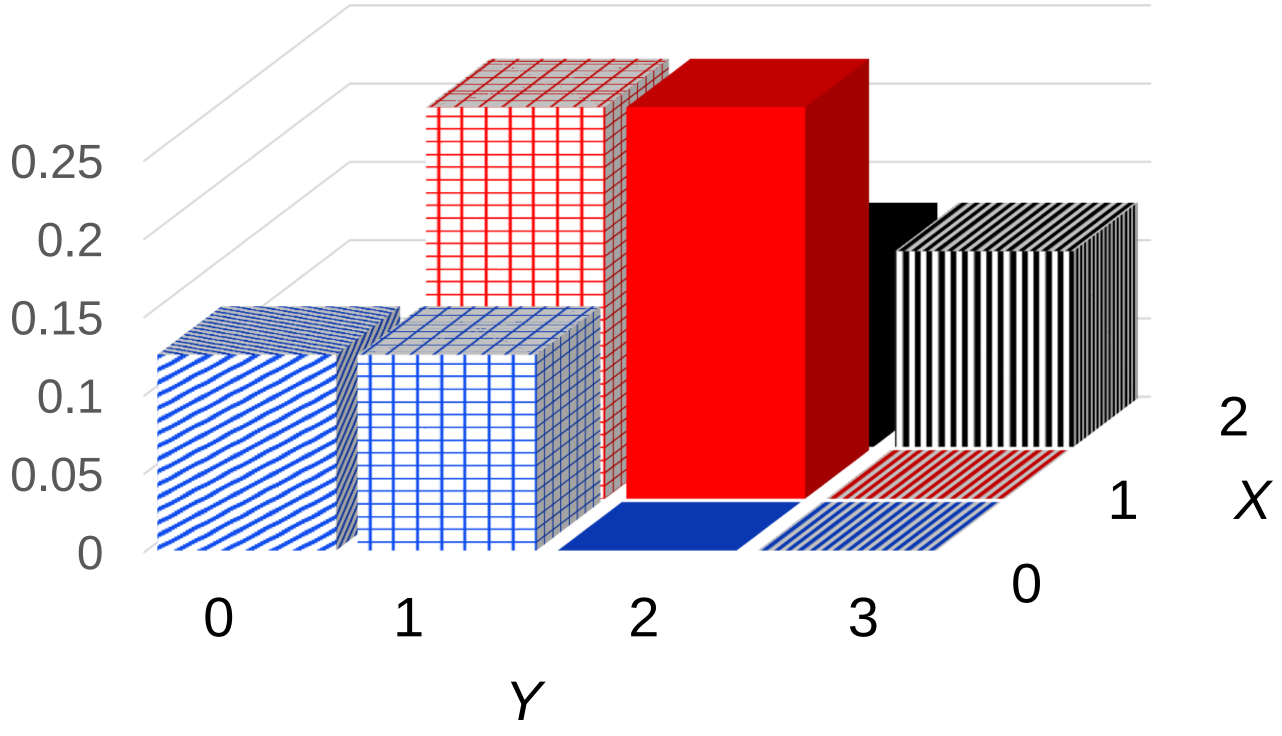


Conditional Distributions

- The **conditional distribution** of *Y* given *X* is defined as

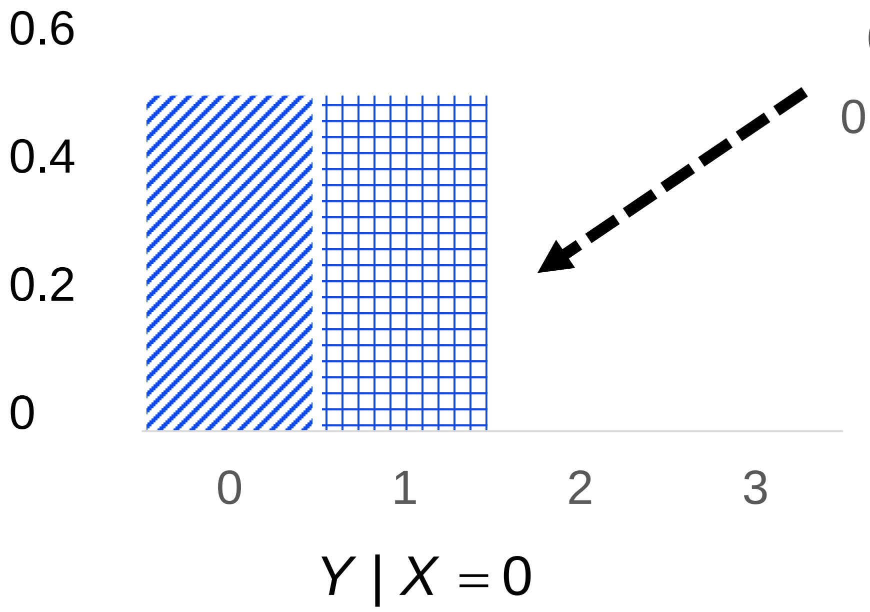
$$P(Y = n \mid X = m) = \frac{P((X,Y) = (m,n))}{P(X = m)}$$

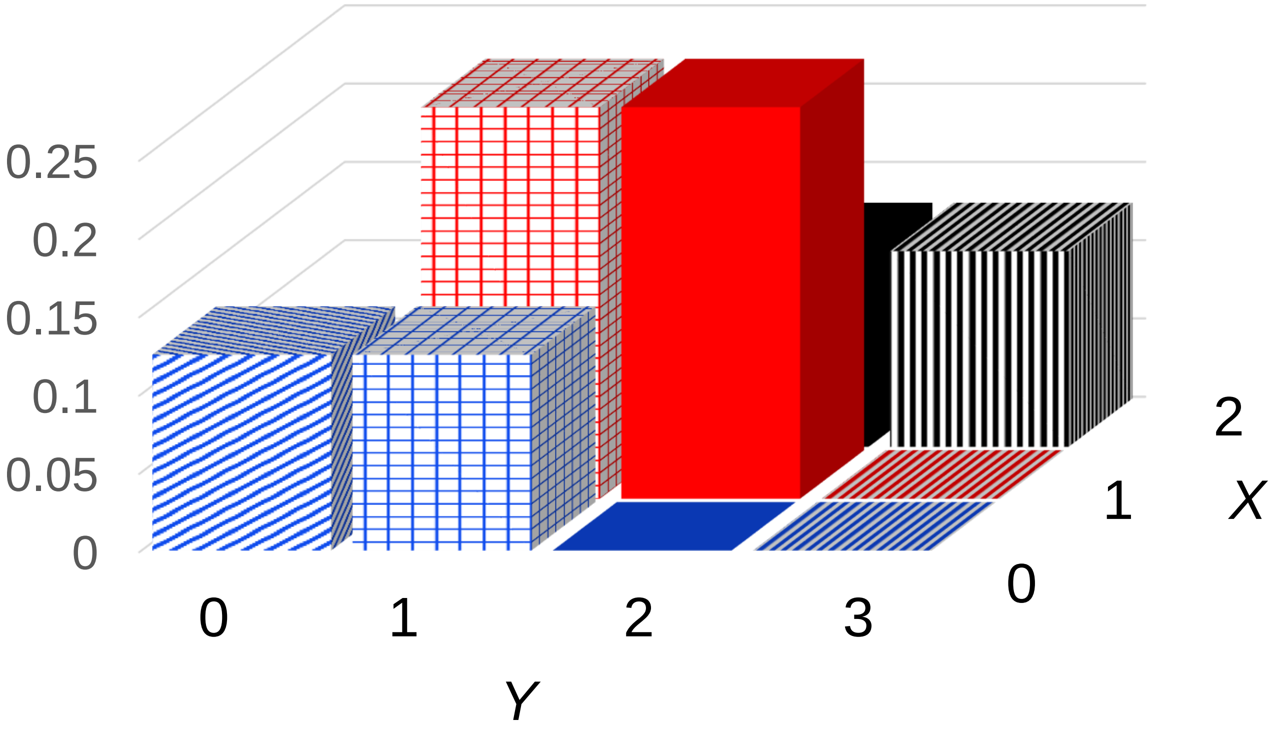
- We can visualise this by taking a slice through the joint density function at a fixed *X* value and renormalising



Conditional Distributions

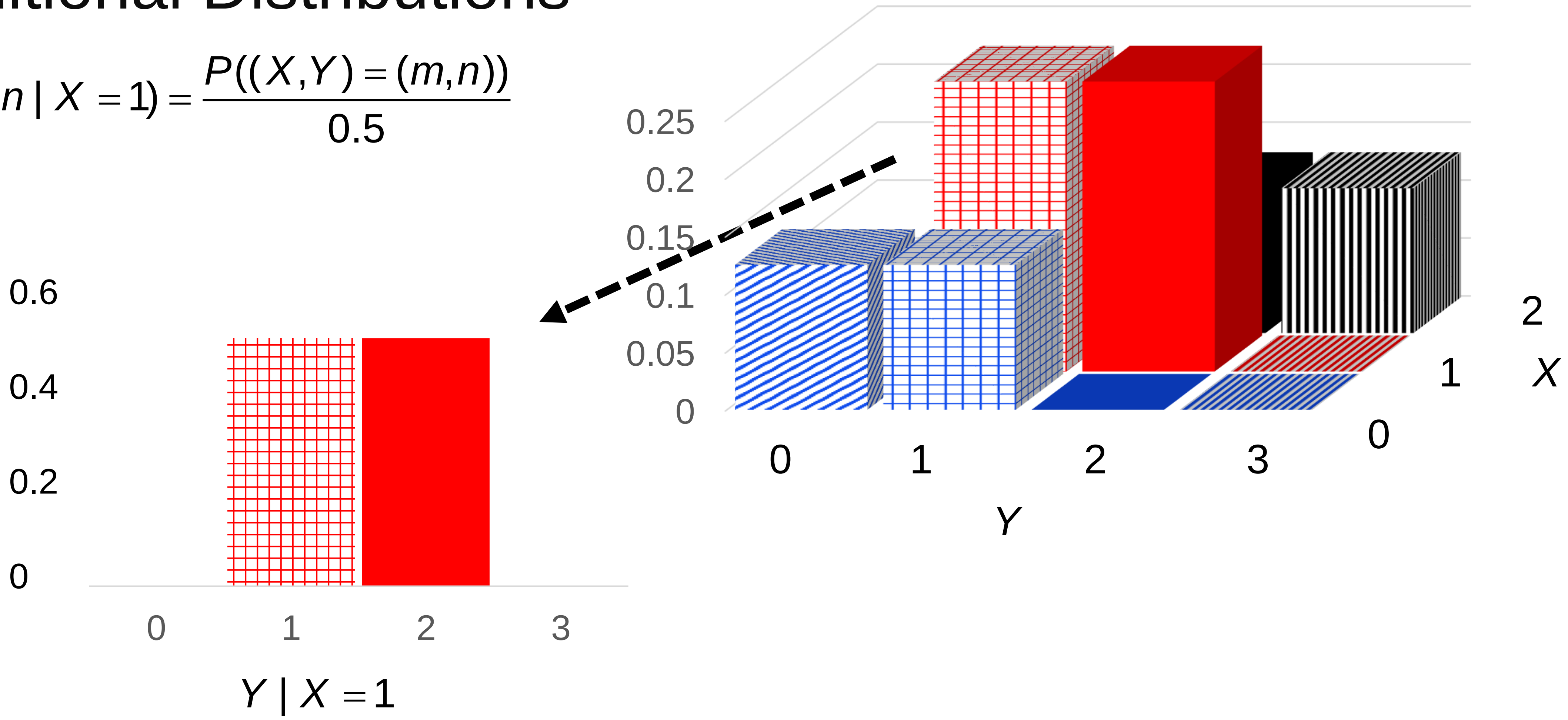
- $P(Y=n \mid X=0)=\dfrac{P((X,Y)=(m,n))}{0.25}$





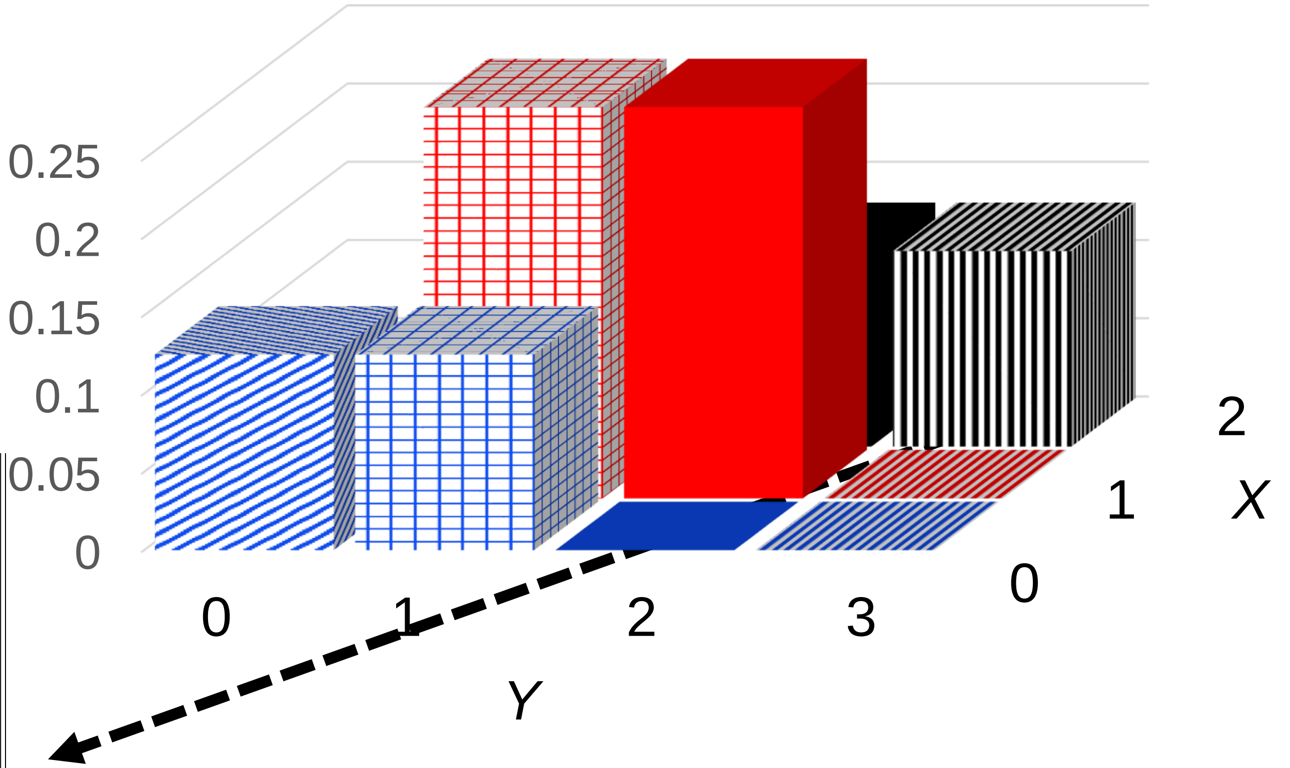
Conditional Distributions

- $P(Y=n \mid X=1)=\dfrac{P((X,Y)=(m,n))}{0.5}$



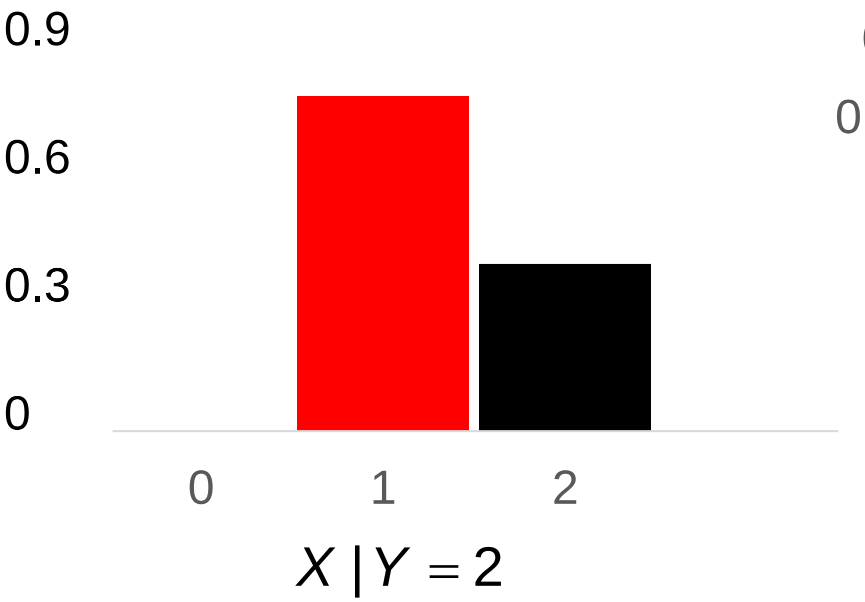
Conditional Distributions

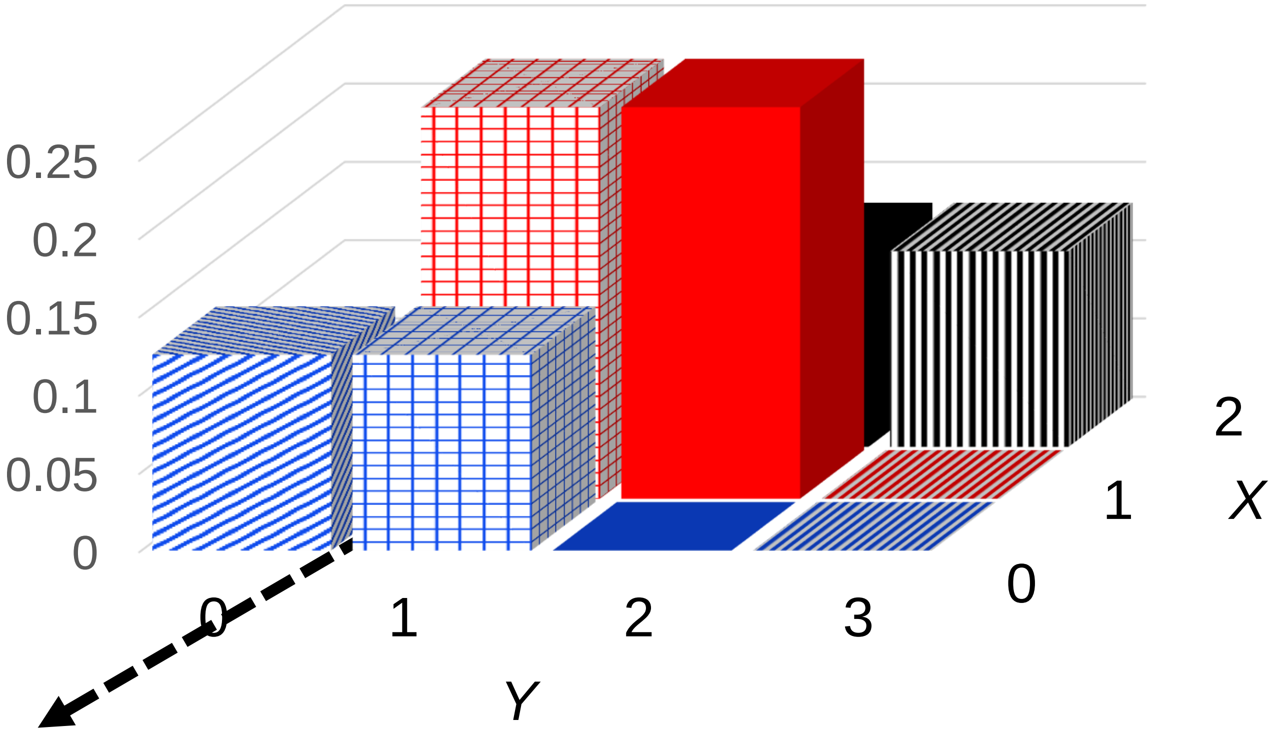
- $P(Y=n \mid X=2) = \dfrac{P((X,Y)=(m,n))}{0.25}$



Conditional Distributions

- $P(X = m \mid Y = 2) = \dfrac{P((X,Y) = (m,n))}{0.375}$





Marginal Distributions

- The **marginal distribution** of *X*, from a joint probability mass function of *X* and *Y* is given by

$$P(X = m) = \sum_{Y} P((X,Y) = (m,n))$$

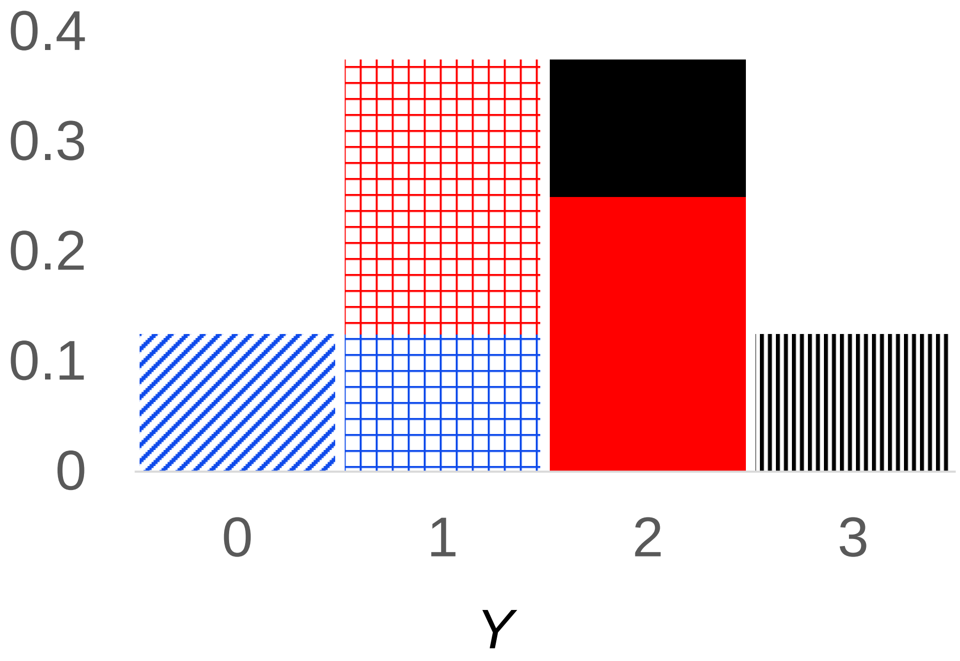
- Unlike the marginal distribution, which assumes knowledge of one (or more) variable, this "averages over" our uncertainties in other variables.

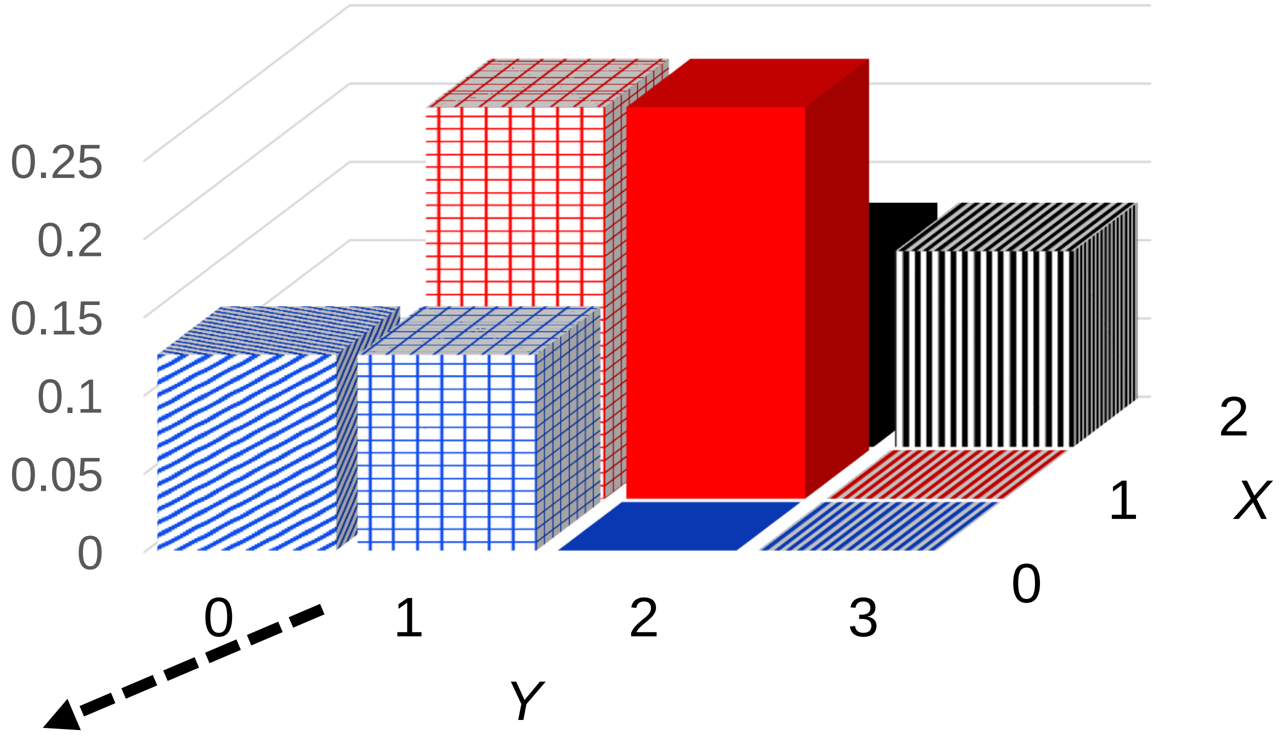
- The marginal distribution there will have variance no lower than the conditional distribution.

- This can also be seen from the Law of Total Variance $Var(Y) = E(Var(Y \mid X)) + Var(E(Y \mid X))$

Marginal Distributions

- We can visualise the marginal distribution by summing all probability masses along a given axis.





Joint Density Functions

- We can extend the idea of joint probability functions to **joint probability density functions** for continuous variables.

- For continuous variables *X* and *Y*, we can define $F(x,y) = P(X \le x \cap Y \le y)$ and hence we have a joint probability density function $f(x,y) = \dfrac{\partial^2 F(x,y)}{\partial x \partial y}$

- As with univariate probability density functions, we have that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\ dxdy = 1$ and $P(X \in [a,b] \cap Y \in [c,d]) = \int_{c}^{d}\int_{a}^{b} f(x,y)\ dxdy$

Marginal and Conditional Density Functions

- Extending the ideas from discrete variables, we can define the **marginal density function** of X as $f(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$.

- The **conditional density** of Y given X is then $f(y\,|\,x)=\dfrac{f(x,y)}{f(x)}$

- Both of these definitions extend to the joint densities of more than three continuous variables as well e.g. $f(x)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y,z)\,dydz$ and $f(x,y\,|\,z)=\dfrac{f(x,y,z)}{f(z)}$

Example

- Consider variables X and Y with joint density function $f(x,y)=\begin{cases}2x^2e^{xy} & (x,y)\in[0,1]^2\\ 0 & \text{otherwise}\end{cases}$

- We calculate the marginal density of X by integrating over all possible values for Y.

$$f(x)=\int_0^1 f(x,y)dy = 2x^2\int_0^1 e^{xy}dy = 2x^2\left[\frac{e^{xy}}{x}\right]_0^1 = 2x^2\left(\frac{e^x-1}{x}\right) = 2x\left(e^x-1\right)$$

- We can verify that this is indeed a valid density function by integrating (by parts) to show that $\int_0^1 f(x)dx = \int_0^1 2x(e^x-1)dx = \left[2x(e^x-x)\right]_0^1 - \int_0^1 2(e^x-x)dx$

$$= \left[2x(e^x-x)\right]_0^1 - \left[2e^x-x^2\right]_0^1 = 2(e-1)-(2e-1-2)=1$$

Example

- This also gives us that the conditional density of Y given X is

$$f(y \mid x) = \frac{f(x,y)}{f(x)} = \frac{2x^2 e^{xy}}{2x\left(e^x - 1\right)} = \frac{xe^{xy}}{\left(e^x - 1\right)}$$

Independence

- Two (or more) random variables are **independent** if and only if the joint probability density or probability mass function factorises into separate functions for each variable.

- For independent discrete variables *X* and *Y*, we have $P((X,Y)=(m,n))=P(X=m)P(Y=n)$ for all *m* and *n*.

- For independent continuous variables *X* and *Y* we have that $f_{X,Y}(x,y)=f_X(x)f_Y(y)$.

- For independent variables, the marginal and conditional distributions are the same e.g. $f(x)=f(x\,|\,y)$

Distribution of a Function of a Random Variable

- For the discrete case, obtaining the probability mass function for a function of a discrete random variable is quite straightforward.

- For example, consider X with mass function $P(X=k)=\begin{cases} 0.3 & k=1 \\ 0.7 & k=3 \\ 0 & \text{otherwise} \end{cases}$ and the distribution of X^2+1.

- As X is discrete, X^2+1 is also discrete.

- X can only take the values 1 and 3 hence X^2+1 can only take the values 2 and 10.

Distribution of a Function of a Random Variable

- $P(X=k)=\begin{cases}0.3 & k=1\\ 0.7 & k=3\\ 0 & \text{otherwise}\end{cases}$ implies that $P(X^2+1=k)=\begin{cases}0.3 & k=2\\ 0.7 & k=10\\ 0 & \text{otherwise}\end{cases}$

- Note that, when the function transforming the variable is not 1-to-1, we sometimes have to combine masses.

- For example $P(Y=k)=\begin{cases}0.3 & k=1\\ 0.3 & k=-1\\ 0.4 & k=3\\ 0 & \text{otherwise}\end{cases}$ implies that $P(Y^2+1=k)=\begin{cases}0.6 & k=2\\ 0.4 & k=10\\ 0 & \text{otherwise}\end{cases}$

Distribution of a Function of a Random Variable

- The situation was more complicated for a continuous random variable.

- If we consider integration by substitution, we have that, given a definite integral

$I = \int_a^b f(x)\,dx$ and a continuous differentiable function $y(x)$ then $I = \int_a^b f(x)\,dx = \int_{y(a)}^{y(b)} f(y)\frac{dx}{dy}\,dy$

- This gives that the density function of $Y(X)$ is given by $g(y) = f(x(y))x'(y)$.

- Note, this assumes that f is one-to-one and $y(b) > y(a)$, but similar statements hold when these assumptions are not met.

Distribution of a Function of Random Variables

- We can extend these ideas to two (or more) variables.
- Given continuous random variables X and Y and a continuous invertible function g such that $(X,Y) = g(S,T)$ then the joint distribution of S and T is given by $f_{S,T}(s,t) = f_{X,Y}(g_1(s,t), g_2(s,t))\left|\det \boldsymbol{J}\right|$ where g_1 and g_2 are the two spatial components of g

and $\boldsymbol{J}$ is the **Jacobian** matrix of this transformation, given by $\boldsymbol{J} = \begin{pmatrix} \dfrac{\partial g_1}{\partial s} & \dfrac{\partial g_2}{\partial s} \\ \dfrac{\partial g_1}{\partial t} & \dfrac{\partial g_2}{\partial t} \end{pmatrix}$

Distribution of a Function of Random Variables

- Let X and Y be independent uniform random variables, $X \sim U[0,1]$ and $Y \sim U[0,1]$.
- $f_X(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$ and $f_Y(y) = \begin{cases} 1 & y \in [0,1] \\ 0 & \text{otherwise} \end{cases}$.
- Since they are independent, the joint density function is just the product of the two density functions hence $f_{X,Y}(x,y) = \begin{cases} 1 & (x,y) \in [0,1]^2 \\ 0 & \text{otherwise} \end{cases}$
- We can calculate the distribution of $T = XY$.

Distribution of a Function of Random Variables

- Inverting $S = X$ and $T = XY$ to make X and Y the subjects, we obtain $X = S$ and $Y = \dfrac{T}{S}$.

- We know the distributions of $X = S$ and $Y = \dfrac{T}{S}$.

- The Jacobian is $\boldsymbol{J} = \begin{pmatrix} \dfrac{\partial}{\partial S}(S) & \dfrac{\partial}{\partial S}\left(\dfrac{T}{S}\right) \\ \dfrac{\partial}{\partial T}(S) & \dfrac{\partial}{\partial T}\left(\dfrac{T}{S}\right) \end{pmatrix} = \begin{pmatrix} 1 & -\dfrac{T}{S^2} \\ 0 & \dfrac{1}{S} \end{pmatrix}$

- The determinant of the Jacobian is therefore $\det(\boldsymbol{J}) = \dfrac{1}{S}$.

Distribution of a Function of Random Variables

- We now wish to calculate $f_T(t) = \int_{-\infty}^{\infty} f_{S,T}(s,t)\, ds = \int_{-\infty}^{\infty} f_{X,Y}(g_1(s,t), g_2(s,t)) \left|\det \boldsymbol{J}\right|\, ds.$

- Since $Y = \dfrac{T}{S}$ and $0 \le Y \le 1$ we have that the density function is only non-zero when $T \le S \le 1.$

- $f_T(t) = \int_{-\infty}^{\infty} f_{X,Y}(g_1(s,t), g_2(s,t)) \left|\det \boldsymbol{J}\right|\, ds = \int_{t}^{1} 1 \times 1 \times \dfrac{1}{s}\, ds \quad = \left[\ln(s)\right]_t^1 = -\ln(t)$

- More fully, we have that $f_T(t) = \begin{cases} -\ln(t) & t \in (0,1] \\ 0 & \text{otherwise} \end{cases}$

Gamma Distribution

- Just as we can extend the idea of a single Bernoulli variable to binomial variables by summing independent Bernoulli variables, we can sum exponential random variables to obtain **gamma variables**.

- That is, if we have $W_1, W_2, W_3, \ldots$ each $\sim \exp(\beta)$, then for integers $\alpha > 0$.

$$S = \sum_{i=1}^{\alpha} W_i \sim Gamma(\alpha, \beta)$$

- The density function of S is given by $f(s) = \begin{cases} \dfrac{\beta^{\alpha} s^{\alpha-1} e^{-\beta s}}{(\alpha-1)!} & s \in [0, \infty) \\ 0 & \text{otherwise} \end{cases}$

- Note that $S \sim Gamma(1, \beta)$ and $S \sim \exp(\beta)$ are equivalent.

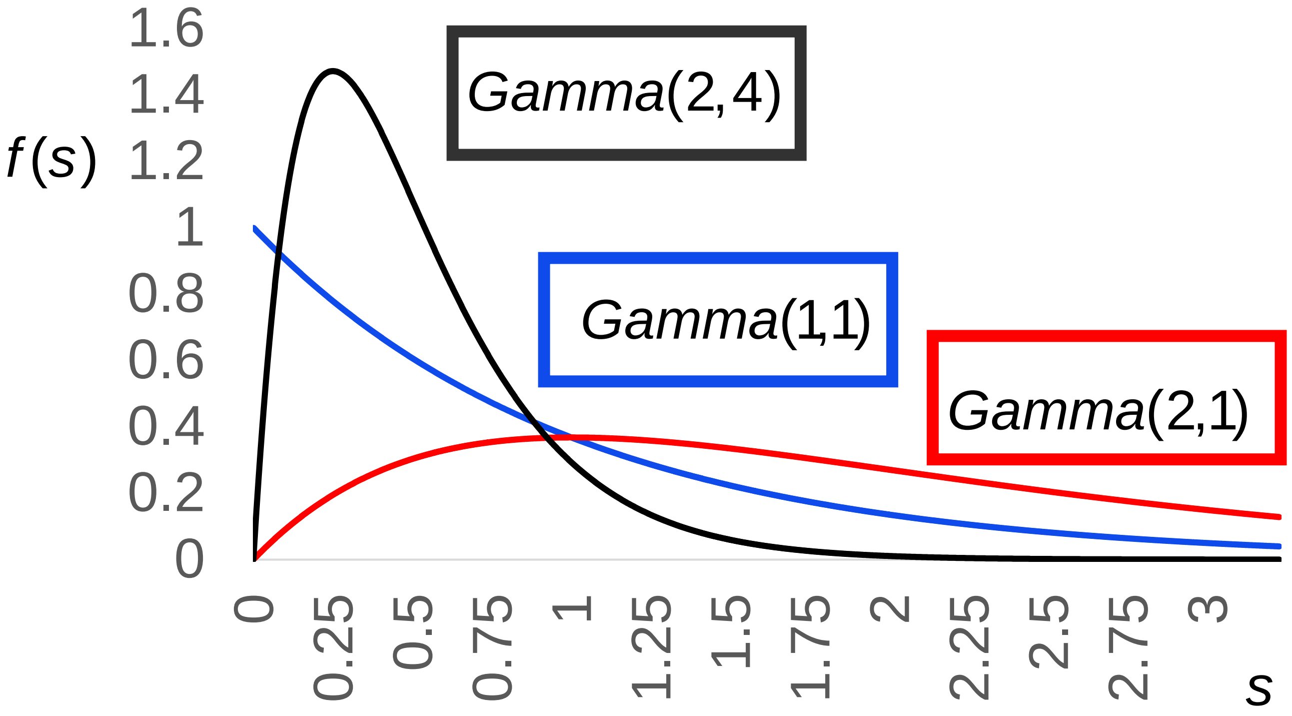
Gamma Distribution

- Depending on the parameters, the gamma distribution can have very different shapes.

- Additionally, the definition of a gamma distribution can be extended to non-integer α

- $f(s) = \begin{cases} \dfrac{\beta^{\alpha} s^{\alpha-1} e^{-\beta s}}{\Gamma(\alpha)} & s \in [0, \infty) \\ 0 & \text{otherwise} \end{cases}$

where $\Gamma(\alpha) = \int_0^{\infty} s^{\alpha-1} e^{-\beta s} ds = (\alpha - 1)!$ when α is a positive integer.



Distributions of Functions of Gamma Variables

- Consider two independent gamma variables, $X \sim gamma(\alpha_X, 1)$ and $Y \sim gamma(\alpha_Y, 1)$.
- We can calculate the distributions of $S = X + Y$ and $R = \dfrac{X}{X+Y}$.
- Making X and Y the subjects of these equations, we obtain $X = RS$ and $Y = (1-R)S$.
- The Jacobian here is $\boldsymbol{J} = \begin{pmatrix} \dfrac{\partial}{\partial S}(RS) & \dfrac{\partial}{\partial S}((1-R)S) \\ \dfrac{\partial}{\partial R}(RS) & \dfrac{\partial}{\partial R}((1-R)S) \end{pmatrix} = \begin{pmatrix} R & (1-R) \\ S & -S \end{pmatrix}$
- This gives $|\det \boldsymbol{J}| = S$.

Distributions of Functions of Gamma Variables

- The joint density function of X and Y is therefore $f_{X,Y}(x,y) = \begin{cases} \dfrac{x^{\alpha_x - 1} e^{-x}}{\Gamma(\alpha_x)} \dfrac{y^{\alpha_y - 1} e^{-y}}{\Gamma(\alpha_y)} & (x,y) \in [0,\infty)^2 \\ 0 & \text{otherwise} \end{cases}$
- This gives $f_{S,R}(s,r) = \begin{cases} \dfrac{(rs)^{\alpha_x - 1} e^{-rs}}{\Gamma(\alpha_x)} \dfrac{((1-r)s)^{\alpha_y - 1} e^{-(1-r)s}}{\Gamma(\alpha_y)} s & (s,r) \in [0,\infty) \times [0,1] \\ 0 & \text{otherwise} \end{cases}$
- Rearranging , we see $f_{S,R}(s,r) = \begin{cases} \left[\dfrac{s^{(\alpha_x + \alpha_y) - 1} e^{-s}}{\Gamma(\alpha_x + \alpha_y)} \right] \left[\dfrac{\Gamma(\alpha_x + \alpha_y)}{\Gamma(\alpha_x)\Gamma(\alpha_y)} r^{\alpha_x - 1} (1-r)^{\alpha_y - 1} \right] & (s,r) \in [0,\infty) \times [0,1] \\ 0 & \text{otherwise} \end{cases}$

Distributions of Functions of Gamma Variables

- $f_{S,R}(s,r)=\begin{cases}\left[\dfrac{s^{(\alpha_x+\alpha_y)-1}e^{-s}}{\Gamma(\alpha_x+\alpha_y)}\right]\left[\dfrac{\Gamma(\alpha_x+\alpha_y)}{\Gamma(\alpha_x)\Gamma(\alpha_y)}r^{\alpha_x-1}(1-r)^{\alpha_y-1}\right] & (s,r)\in[0,\infty)\times[0,1] \\ 0 & \text{otherwise}\end{cases}$ factorises into a

function in *S* which does not depend on *R* and one in *R* which does not depend on *S*.

- *R* and *S* are therefore independent.

- Furthermore, they each have a standard distribution.

Distributions of Functions of Gamma Variables

- $$f_{S,R}(s,r) = \begin{cases} \left[\dfrac{s^{(\alpha_x+\alpha_y)-1}e^{-s}}{\Gamma(\alpha_x+\alpha_y)}\right]\left[\dfrac{\Gamma(\alpha_x+\alpha_y)}{\Gamma(\alpha_x)\Gamma(\alpha_y)} r^{\alpha_x-1}(1-r)^{\alpha_y-1}\right] & (s,r)\in[0,\infty)\times[0,1] \\ 0 & \text{otherwise} \end{cases}$$

- $f_{S,R}(s,r) = f_S(s)f_R(r)$ where $f_S(s) = \begin{cases} \left[\dfrac{s^{(\alpha_x+\alpha_y)-1}e^{-s}}{\Gamma(\alpha_x+\alpha_y)}\right] & s\in[0,\infty) \\ 0 & \text{otherwise} \end{cases}$

  and $f_R(r) = \begin{cases} \left[\dfrac{\Gamma(\alpha_x+\alpha_y)}{\Gamma(\alpha_x)\Gamma(\alpha_y)} r^{\alpha_x-1}(1-r)^{\alpha_y-1}\right] & r\in[0,1] \\ 0 & \text{otherwise} \end{cases}$

Distributions of Functions of Gamma Variables

- $$f_S(s) = \begin{cases} \left[\dfrac{s^{(\alpha_x+\alpha_y)-1}e^{-s}}{\Gamma(\alpha_x+\alpha_y)}\right] & s \in [0,\infty) \\ 0 & \text{otherwise} \end{cases}$$ hence $S \sim gamma(\alpha_x+\alpha_y, 1)$.

- $$f_R(r) = \begin{cases} \left[\dfrac{\Gamma(\alpha_x+\alpha_y)}{\Gamma(\alpha_x)\Gamma(\alpha_y)} r^{\alpha_x-1}(1-r)^{\alpha_y-1}\right] & r \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$ hence $R \sim Beta(\alpha_x, \alpha_y)$.

- We will see more of the **beta distribution** later in the subject in the context of Bayesian statistics.
- Here, we have proven that it can be obtained by taking the ratio one gamma variable to the sum of itself and an independent gamma variable.