

Probability Theory

Chapter 6—Jointly Distributed Random Variables

Nan Liu¹

¹These slides are based on Prof. Mui Li's. Any mistakes are my own. If you find an error, please send an email at nanliu@xmu.edu.cn

6.1 Jointly Distributed Random Variables

- Assume we have two r.v. X and Y , then we define the joint c.d.f.

$$F(a, b) = P(X \leq a, Y \leq b)$$

- The c.d.f of X is

$$\begin{aligned} F_X(a) &= P(X \leq a, Y \leq \infty) \\ &= \lim_{b \rightarrow \infty} P(X \leq a, Y \leq b) \\ &= \lim_{b \rightarrow \infty} F(a, b) \end{aligned}$$

- Similarly we have that the c.d.f. of Y is

$$F_Y(b) = P(X \leq \infty, Y \leq b) = \lim_{a \rightarrow \infty} F(a, b)$$

6.1 Jointly Distributed Random Variables

- Consider

$$\begin{aligned}P(X > a, Y > b) &= 1 - P(\{X > a, Y > b\}^c) \\&= 1 - P(\{X > a, Y > b\}^c) \\&= 1 - P(\{X > a\}^c \cup \{Y > b\}^c) \\&= 1 - P(\{X \leq a\} \cup \{Y \leq b\}) \\&= 1 - [P(X \leq a) + P(Y \leq b) - P(X \leq a, Y \leq b)] \\&= 1 - F_X(a) - F_Y(b) + F(a, b).\end{aligned}$$

6.1 Probability mass Functions – Discrete R.V.s

- For discrete variables, we work directly with the joint pmf

$$p(x, y) = P(X = x, Y = y)$$

from which we obtain the marginal pmf

$$p_X(x) = \sum_y p(x, y), \quad p_Y(y) = \sum_x p(x, y).$$

6.1 Example

- Suppose that 3 balls are randomly selected from an urn containing 3 red, 4 white, and 5 blue balls. If we let X and Y denote, respectively, the number of red and white balls chosen, then the joint probability mass function of X and Y , $p(i, j) = P\{X = i, Y = j\}$, is given by

$i \backslash j$					
	0	1	2	3	Row sum = $P\{X = i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Column sum = $P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

6.1 Joint Probability Density Function–Continuous R.V.s

- If both X and Y are jointly continuous, then their joint p.d.f. is a non-negative function $f(x, y)$ such that for any set C

$$P\{(X, Y) \in C\} = \int \int_{(x,y) \in C} f(x, y) \, dx dy.$$

- In particular, we have

$$F(a, b) = P(X \leq a, Y \leq b) = \int_{-\infty}^a \int_{-\infty}^b f(x, y) \, dx dy.$$

- Hence, when we differentiate, we obtain

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y).$$

6.1 Marginal Probability Density Functions—Continuous R.V.s

- The marginal pdf f_X of the r.v. X is defined by

$$F(a) = P(X \leq a) = \int_{-\infty}^a f_X(x) dx$$

- We have

$$P(X \leq a) = P(X \leq a, Y \leq \infty) = \int_{-\infty}^a \int_{-\infty}^{\infty} f(x, y) dx dy$$

so clearly

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy.$$

- Similarly, we have

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

6.1 Example

- Consider the joint p.d.f.

$$f(x, y) = \begin{cases} kx & \text{for } 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant. Find k ?

- We have $\int_0^1 \int_0^1 f(x, y) dx dy = 1$ where

$$\begin{aligned} \int_0^1 \int_0^1 f(x, y) dx dy &= \int_0^1 \int_0^1 kx dy dx = \int_0^1 kx dx \\ &= \left[k \frac{x^2}{2} \right]_0^1 = \frac{k}{2} = 1 \end{aligned}$$

thus $k = 2$.

6.1 Example: System Failure

- A device contains two components. The device fails if either component fails. The joint p.d.f. of the lifetimes of the components, measured in hours, is $f(s, t)$ where $0 < s < 1$ and $0 < t < 1$. What is the probability that the device fails during the first half hour of operation?
- We have

$$\begin{aligned}\text{Proba} &= P[(S \leq 0.5) \cup (T \leq 0.5)] \\ &= \int_0^{0.5} \int_{0.5}^1 f(s, t) \, ds dt + \int_0^1 \int_0^{0.5} f(s, t) \, ds dt\end{aligned}$$

as this follows directly from a graph.

6.1 Example: System Failure

- A device runs until either of two components fails, at which point the device stops running. The joint density function of the lifetimes of the two components, both measured in hours, is

$$f(x, y) = \frac{x+y}{8} \text{ for } 0 < x < 2 \text{ and } 0 < y < 2.$$

What is the probability that the device fails during its first hour of operation?

- We have

$$\begin{aligned} \text{Prob.} &= P(\{X \leq 1\} \cup \{Y \leq 1\}) \\ &= P([\{X > 1\} \cap \{Y > 1\}]^c) \\ &= 1 - P([\{X > 1\} \cap \{Y > 1\}]) \end{aligned}$$

where

$$P([\{X > 1\} \cap \{Y > 1\}]) = \int_1^2 \int_1^2 \frac{x+y}{8} dx dy.$$

6.1 Example: System Failure

- We have

$$\begin{aligned}\frac{1}{8} \int_1^2 \int_1^2 (x+y) \, dx dy &= \frac{1}{8} \int_1^2 \left[\frac{(x+y)^2}{2} \right]_1^2 dy \\&= \frac{1}{18} \int_1^2 \left[(y+2)^2 - (y+1)^2 \right] dy \\&= \frac{1}{18} \left(\left[\frac{(y+2)^3}{3} \right]_1^2 - \left[\frac{(y+1)^3}{3} \right]_1^2 \right) = \frac{18}{48}\end{aligned}$$

- Hence we have Prob. = $1 - \frac{18}{48} = \frac{30}{48} = 0.625$.

6.1 Example: Insurance Company

- An insurance company insures a large number of drivers. Let X be the r.v. representing the company's losses under collision insurance, and let Y represent the company's losses under liability insurance. X and Y have a joint p.d.f

$$f(x, y) = \begin{cases} \frac{2x+2-y}{4} & \text{for } 0 < x < 1 \text{ and } 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

What is the probability that the total loss is at least 1?

- We need to compute

$$\begin{aligned} P(X + Y > 1) &= \int \int_{(x,y): x+y>1} f(x, y) \, dx dy \\ &= \int_0^1 \left[\int_{1-x}^2 f(x, y) \, dy \right] dx \\ &= \int_0^1 \left[\left(\frac{x+1}{2} \right) y - \frac{y^2}{8} \right]_{1-x}^2 dx \end{aligned}$$

6.1 Example: Insurance Company

- We have

$$\begin{aligned}P(X + Y > 1) &= \int_0^1 \left[\frac{5}{8}x^2 + \frac{3}{4}x + \frac{1}{8} \right] dx \\&= \int_0^1 \left(\frac{5}{8}x^2 + \frac{3}{4}x + \frac{1}{8} \right) dx \\&= \left[\frac{5}{24}x^3 + \frac{3}{8}x^2 + \frac{x}{8} \right]_0^1 \\&= \frac{5}{24} + \frac{3}{8} + \frac{1}{8} = \frac{17}{24}.\end{aligned}$$

6.2 Independent Random Variables

- The r.v. X and Y are said to be independent if, for any sets A and B we have

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$$

- It can be shown that X and Y are independent if and only if

$$F(x, y) = F_X(x) F_Y(y)$$

or

$$f(x, y) = f_X(x) f_Y(y)$$

or

$$f_{X,Y}(x, y) = h(x)g(y) \quad -\infty < x < \infty, -\infty < y < \infty$$

6.2 Example

- Suppose that $n + m$ independent trials having a common probability of success p are performed. If X is the number of successes in the first n trials, and Y is the number of successes in the final m trials, then X and Y are independent, since knowing the number of successes in the first n trials does not affect the distribution of the number of successes in the final m trials (by the assumption of independent trials).
- In fact, for integral x and y ,

$$\begin{aligned}P\{X = x, Y = y\} &= \binom{n}{x} p^x (1 - p)^{n-x} \binom{m}{y} p^y (1 - p)^{m-y} \\&= P\{X = x\} P\{Y = y\}\end{aligned}$$

- In contrast, X and Z will be dependent, where Z is the total number of successes in the $n + m$ trials. (Why?)

6.2 Example

- The joint density of two r.v. X and Y is given by

$$f(x, y) = \begin{cases} x e^{-(x+y)} & \text{for } x > 0 \text{ and } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Are the r.v. X and Y independent?

- We need to check whether $f(x, y) = f_X(x) f_Y(y)$ or not. We have for $x > 0$

$$f_X(x) = \int_0^{\infty} f(x, y) dy = x e^{-x} \int_0^{\infty} e^{-y} dy = x e^{-x}$$

and for $y > 0$

$$\begin{aligned} f_Y(y) &= \int_0^{\infty} f(x, y) dx = e^{-y} \int_0^{\infty} x e^{-x} dx \\ &= e^{-y} \left\{ [-x e^{-x}]_0^{\infty} + \int_0^{\infty} e^{-x} dx \right\} \\ &= e^{-y} \end{aligned}$$

so X and Y are independent rv.

6.2 Example: Insurance Policies

- An insurance company sells two types of auto insurance policies: Basic and Deluxe. The time until the next Basic policy claim is an exponential random variable with mean two days. The time until the next Deluxe policy claim is an independent exponential random variable with mean three days. What is the probability that the next claim will be a Deluxe policy claim?
- In this case, we have for $0 < t_1 < \infty$, $0 < t_2 < \infty$

$$f(t_1, t_2) = \left(\frac{1}{2} \exp(-t_1/2) \right) \left(\frac{1}{3} \exp(-t_2/3) \right)$$

and we want to find

$$P(T_2 < T_1) = \int_0^\infty \int_0^{t_1} f(t_1, t_2) dt_1 dt_2.$$

6.2 Example: Insurance Policies

- We have

$$\begin{aligned}P(T_2 < T_1) &= \int_0^\infty \left[\int_0^{t_1} f(t_1, t_2) dt_2 \right] dt_1 \\&= \int_0^\infty \exp(-t_1/2) \left[-\frac{1}{2} \exp(-t_2/3) \right]_0^{t_1} dt_1 \\&= \int_0^\infty \left(\frac{1}{2} \exp(-t_1/2) - \frac{1}{2} \exp(-t_1/2) \exp(-t_1/3) \right) dt_1 \\&= \int_0^\infty \left(\frac{1}{2} \exp(-t_1/2) - \frac{1}{2} \exp(-5t_1/6) \right) dt_1 \\&= \left[-\exp(-t_1/2) + \frac{3}{5} \exp(-5t_1/6) \right]_0^\infty = 1 - \frac{3}{5} = \frac{2}{5}.\end{aligned}$$

6.2 Example: Operational Cost

- A device containing two key components fails when, and only when, both components fail. The lifetimes, T_1 and T_2 , of these components are independent with common density function $f(t) = \exp(-t)$ for $t \geq 0$. The cost X of operating the device until failure is $2T_1 + T_2$. What is the density function of X ?
- We have $f_X(x) > 0$ only if $x \geq 0$. For $x \geq 0$, we have

$$\begin{aligned}P(X \leq x) &= P(2T_1 + T_2 \leq x) \\&= \int_0^x \exp(-t_2) \left[\int_0^{\frac{1}{2}(x-t_2)} \exp(-t_1) dt_1 \right] dt_2 \\&= \int_0^x \exp(-t_2) [-\exp(-t_1)]_0^{\frac{1}{2}(x-t_2)} dt_2 \\&= \int_0^x \exp(-t_2) \left[1 - \exp\left(-\frac{1}{2}(x-t_2)\right) \right] dt_2 \\&= \left[-\exp(-t_2) + 2 \exp\left(-\frac{x}{2}\right) \exp\left(-\frac{t_2}{2}\right) \right]_0^x\end{aligned}$$

6.2 Example: Operational Cost

- We have

$$\begin{aligned}P(X \leq x) &= \left[-\exp(-t_2) + 2 \exp\left(-\frac{x}{2}\right) \exp\left(-\frac{t_2}{2}\right) \right]_0^x \\&= 1 - 2 \exp\left(-\frac{x}{2}\right) + \exp(-x).\end{aligned}$$

- Hence, the density is given by

$$\begin{aligned}f_X(x) &= \frac{dP(X \leq x)}{dx} \\&= \exp\left(-\frac{x}{2}\right) - \exp(-x)\end{aligned}$$

for $x \geq 0$.

6.3 Sum of Independent Random Variables–Discrete

- Consider two integer-valued independent r.v. X and Y of respective p.m.f. $p_X(x)$ and $p_Y(y)$.
- Consider $Z = X + Y$, we want to compute the p.m.f. of Z denoted $p_Z(z)$.
- Assume $Y = y$ then $Z = z$ if and only if $X = z - y$ and

$$P(X = z - y \cap Y = y) = p_X(z - y) p_Y(y)$$

so, as Y can take integer values and the events $(X = z - y) \cap (Y = y)$ and $(X = z - y') \cap (Y = y')$ are mutually exclusive for $y \neq y'$, we have

$$p_Z(z) = \sum_{y=-\infty}^{\infty} p_X(z - y) p_Y(y).$$

6.3 Sum of Poisson Random Variables

- Consider two Poisson independent r.v. X and Y of respective p.m.f. $p_X(x) = e^{-\lambda} \frac{\lambda^x}{x!}$ for $x = 0, 1, 2, \dots$ and $p_Y(y) = e^{-\lambda'} \frac{\lambda'^y}{y!}$.
- Consider $Z = X + Y$, then we can prove that Z is Poisson of parameter $\lambda + \lambda'$.
- Proof:** For $z \geq 0$ we have $Y = y \leq z$ that $X = z - y$

$$\begin{aligned} p_Z(z) &= e^{-(\lambda+\lambda')} \sum_{y=0}^z \frac{\lambda^{(z-y)}}{(z-y)!} \frac{(\lambda')^y}{y!} \\ &= \frac{e^{-(\lambda+\lambda')}}{z!} \sum_{y=0}^z \binom{z}{y} (\lambda')^y \lambda^{(z-y)} \\ &= \frac{e^{-(\lambda+\lambda')}}{z!} (\lambda + \lambda')^z \quad (\text{by binomial formula}) \end{aligned}$$

6.3 Sum of Independent Random Variables–Continuous

- In numerous scenarios, we have to sum independent continuous r.v.; signal + noise, sums of different random effects etc.
- Assume that X, Y are continuous r.v. of respective pdf $f_X(x)$ and $f_Y(y)$ then $Z = X + Y$ admits the pdf

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx \end{aligned}$$

- The pdf $f_Z(z)$ is the so-called “convolution” of $f_X(x)$ and $f_Y(y)$.

6.3 Proof

- We have

$$\begin{aligned}F_Z(z) &= P(Z \leq z) = P(X + Y \leq z) \\&= \int \int_{(x+y \leq z)} f_X(x) f_Y(y) dx dy \\&= \int \int_{(y \in \mathbb{R}; x \leq z-y)} f_X(x) f_Y(y) dx dy \\&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-y} f_X(x) dx \right) f_Y(y) dy \\&= \int_{-\infty}^{\infty} F_X(z-y) f_Y(y) dy\end{aligned}$$

- Now we obtain

$$\begin{aligned}f_Z(z) &= \frac{dF_Z(z)}{dz} = \int_{-\infty}^{\infty} \frac{dF_X(z-y)}{dz} f_Y(y) dy \\&= \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy \quad (\text{chain rule})\end{aligned}$$

6.3 Sum of Uniform Random Variables

- In this case, we have $f_X(x) = 1_{[0,1]}(x)$ and $f_Y(y) = 1_{[0,1]}(y)$ so $f_Z(z)$ is non-null in $[0, 2]$
- We have

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy \\ &= \int_0^1 f_X(z-y) dy = \int_0^1 1_{[0,1]}(z-y) dy \end{aligned}$$

where

$$\int_0^1 1_{[0,1]}(z-y) dy = \int_{z-1}^z 1_{[0,1]}(u) du = \begin{cases} z & \text{if } 0 \leq z \leq 1 \\ 2-z & \text{if } 1 \leq z \leq 2 \end{cases}$$

6.3 Sum of Exponential Random Variables

- Consider two independent exponential r.v. X, Y of parameter λ (i.e. $f_X(x) = f_Y(x) = \lambda e^{-\lambda x} 1_{[0, \infty)}(x)$).
- The pdf of $Z = X + Y$ is $f_Z(z) = 0$ for $z < 0$ and for $z > 0$

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} \lambda e^{-\lambda(z-y)} 1_{[0, \infty)}(z-y) \lambda e^{-\lambda y} 1_{[0, \infty)}(y) dy \\ &= \lambda^2 e^{-\lambda z} \int_0^{\infty} 1_{[0, \infty)}(z-y) dy \\ &= \lambda^2 e^{-\lambda z} \int_0^z 1_{[0, \infty)}(z-y) dy \\ &= \lambda^2 z e^{-\lambda z}. \end{aligned}$$

6.3 Sum of Gaussian Random Variables

- Consider two independent normal standard r.v. X, Y (i.e. $f_X(x) = f_Y(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$) then the pdf of $Z = X + Y$ is

$$f_Z(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(z-y)^2/2} e^{-y^2/2} dy$$

where

$$\begin{aligned}(z-y)^2 + y^2 &= z^2 + 2y^2 - 2yz \\ &= z^2 + 2(y - z/2)^2 - z^2/2 = 2(y - z/2)^2 + z^2/2\end{aligned}$$

- So we have

$$f_Z(z) = \frac{e^{-z^2/4}}{2\pi} \sqrt{\pi} \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-(y-z/2)^2} dy = \frac{e^{-z^2/4}}{\sqrt{2\pi}\sqrt{2}}$$

Hence Z is a normal r.v. of mean 0 and variance 2.

- Generalization:** if X is an normal r.v. (μ_X, σ_X^2) and Y is an normal r.v. (μ_Y, σ_Y^2) where X and Y are independent then Z is a normal r.v. $(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

6.4 Conditional Distributions: Discrete Case

- Given a joint p.m.f. for two r.v. X, Y it is possible to compute the conditional p.m.f. X given $Y = y$.
- Assume X, Y are discrete-valued r.v. with a joint p.m.f. $p(x, y)$ then the conditional p.m.f. of X given $Y = y$ is

$$\begin{aligned} p_{X|Y}(x|y) &: = P(X = x | Y = y) \\ &= \frac{P(X = x \cap Y = y)}{P(Y = y)} \\ &= \frac{p(x, y)}{p_Y(y)}. \end{aligned}$$

- In the case where X and Y are independent, we have $p_{X|Y}(x|y) = p_X(x)$ as $p(x, y) = p_X(x) p_Y(y)$.

6.4 Conditional Distributions: Discrete Case

- We have

$$p(x, y) = p_{X|Y}(x|y) p_Y(y)$$

and similarly

$$p(x, y) = p_{Y|X}(y|x) p_X(x)$$

- Hence we obtain

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x) p_X(x)}{p_Y(y)}$$

which holds if $p_Y(y) > 0$.

6.4 Conditional Expectation and Variance: Discrete Case

- We can define the mean, variance of the conditional p.m.f.
- The conditional mean is given by

$$E(X|Y = y) = \sum_x x \cdot p_{X|Y}(x|y)$$

- The conditional variance is given by

$$\begin{aligned} \text{Var}(X|Y = y) &= E\left((X - E(X|Y = y))^2 | Y = y\right) \\ &= E(X^2|Y = y) - \{E(X|Y = y)\}^2 \end{aligned}$$

where

$$E(X^2|Y = y) = \sum_x x^2 \cdot p_{X|Y}(x|y)$$

- $E(X|Y = y)$ and $\text{Var}(X|Y = y)$ are functions but $E(X|Y)$ and $\text{Var}(X|Y)$ are random variables.

6.4 Example: Toy problem

- Consider $X \in \{0, 1, 2\}$ and $Y \in \{0, 1, 2\}$ such that their joint pmf is given by

		y		
		0	1	2
x	0	1/9	2/9	1/9
	1	2/9	2/9	0
	2	1/9	0	0

- The conditional pmf of X given $Y = 0$ and $Y = 1$ are

$$p_{X|Y}(x|0) = \frac{p(x, y=0)}{1/9 + 2/9 + 1/9}, \quad p_{X|Y}(x|1) = \frac{p(x, y=1)}{2/9 + 2/9 + 0}.$$

- We have

$$\begin{aligned} E(X|0) &= 1 \times p_{X|Y}(x|0) + 2 \times p_{X|Y}(x|0) \\ &= \frac{2/9}{4/9} + 2 \times \frac{1/9}{4/9} = 1 \end{aligned}$$

6.4 Example: Fair Die

- Roll a die until we get a 6. Let Y be the total number of rolls and X the number of 1's we get. What is the conditional pmf $p_{X|Y}(x|y)$? Compute $E(X|y)$ and $Var(X|Y=y)$.
- The event $Y = y$ means that there were $y - 1$ rolls that were not a 6 and then the y th roll was a six.
- So $p_{X|Y}(x|y)$ is a binomial distribution with $n = y - 1$ trials and proba. of success $p = 1/5$.
- It follows that

$$E(X|y) = np = \frac{(y-1)}{5},$$

$$Var(X|y) = np(1-p) = \frac{4(y-1)}{25}$$

- If we do not observe $Y = y$, then $E(X|Y) = \frac{Y-1}{5}$ and $Var(X|Y) = \frac{4(Y-1)}{25}$ are not numbers but random variables.

6.4 Example: Fair Die

- As $E(X|Y)$ is a random variable, it is possible to compute its expectation

$$\begin{aligned}E(E(X|Y)) &= \sum_y E(X|Y=y) \cdot p_Y(y) \\&= \sum_y \frac{(y-1)}{5} \cdot p_Y(y) \\&= -\frac{1}{5} + \frac{1}{5} \sum_y y p_Y(y) = -\frac{1}{5} + 1/5 * E(Y) \\&= -\frac{1}{5} + 6/5 = 1 \text{ as } Y \sim \text{Geometric}(1/6)\end{aligned}$$

- It can actually be easily established that

$$E(E(X|Y)) = E(X).$$

6.4 Example: Poisson random variables

- Consider two Poisson independent r.v. X and Y of respective p.m.f. $p_X(x) = e^{-\lambda} \frac{\lambda^x}{x!}$ for $x = 0, 1, 2, \dots$ and $p_Y(y) = e^{-\lambda'} \frac{\lambda'^y}{y!}$. Calculate the conditional p.m.f. of X given $X + Y = m$. What is the conditional expectation and variance $E(X|y)$ and $Var(X|y)$?
- We have $p(X = x | X + Y = m) = 0$ for $x > m$ and for $x \leq m$

$$\begin{aligned} p(X = x | X + Y = m) &= \frac{p(X=x, X+Y=m)}{p(X+Y=m)} \\ &= \frac{p(X=x, Y=m-x)}{p(X+Y=m)} = \frac{p_X(X=x) p_Y(Y=m-x)}{p(X+Y=m)} \\ &= \frac{e^{-\lambda} \frac{\lambda^x}{x!} e^{-\lambda'} \frac{(\lambda')^{m-x}}{(m-x)!}}{e^{-(\lambda+\lambda')} \frac{(\lambda+\lambda')^m}{m!}} \text{ as } X + Y \text{ is Poisson } \lambda + \lambda' \\ &= \binom{m}{x} \frac{\lambda^x (\lambda')^{m-x}}{(\lambda+\lambda')^m} = \binom{m}{x} \left(\frac{\lambda}{\lambda+\lambda'} \right)^x \left(1 - \frac{\lambda}{\lambda+\lambda'} \right)^{m-x} \end{aligned}$$

which is a Binomial of parameter $n = m$ and success proba $p = \lambda / (\lambda + \lambda')$.

6.4 Example: Poisson random variables

- Hence we have

$$E(X|X+Y=m) = np = \frac{m\lambda}{(\lambda + \lambda')}$$

- We also obtain

$$\begin{aligned} \text{Var}(X|X+Y=m) &= np(1-p) \\ &= \frac{m\lambda\lambda'}{(\lambda + \lambda')^2} \end{aligned}$$

6.4 Example: How Many Tax Fraudsters?

- The number N of tax fraudsters is assumed to follow a Poisson distribution with param λ . Each tax fraudster is identified with proba p independently of the other fraudsters. Let K be the number of fraudsters identified. What is the conditional p.m.f. of N given $K = k$? Compute $E(N|k)$.
- We have

$$p_N(n) = e^{-\lambda} \frac{\lambda^n}{n!},$$
$$p_{K|N}(k|n) = \binom{n}{k} p^k (1-p)^{n-k}.$$

- We want to compute

$$p_{N|K}(n|k) = \frac{p_{K|N}(k|n) p_N(n)}{p_K(k)}$$

6.4 Example: How Many Tax Fraudsters?

- We have $p_{N|K}(n|k) = 0$ if $n < k$.
- If $n \geq k$

$$\begin{aligned} p_{N|K}(n|k) &= \frac{\binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!}}{\sum_{m \geq k} \binom{m}{k} p^k (1-p)^{m-k} e^{-\lambda} \frac{\lambda^m}{m!}} \\ &= \frac{\{(1-p)\lambda\}^{n-k}}{(n-k)!} e^{-(1-p)\lambda} \quad (\text{use } i = m - k) \end{aligned}$$

- Hence we have

$$E(N|k) = \sum_{n \geq k} n \frac{\{(1-p)\lambda\}^{n-k}}{(n-k)!} e^{-(1-p)\lambda} = k + (1-p)\lambda$$

6.5 Conditional distribution: Continuous Case

- Given a joint p.d.f. for two r.v. X and Y , it is possible to compute the conditional p.d.f. of X having observed $Y = y$.
- Assume X, Y are continuous-valued r.v. with a joint p.d.f. $f(x, y)$, then the conditional pdf of X given $Y = y$ is

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

- This can be heuristically established by noting that

$$\begin{aligned} f_{X|Y}(x|y)dx &= \frac{f(x, y)dx dy}{f_Y(y)dy} \\ &\approx \frac{P(x \leq X \leq x + dx \cap y \leq Y \leq y + dy)}{P(y \leq Y \leq y + dy)} \\ &= P(x \leq X \leq x + dx | y \leq Y \leq y + dy) \end{aligned}$$

- In the case where X and Y are independent, we have,
 $f_{X|Y}(x|y) = f_X(x)$ and $f(x, y) = f(x)f(y)$.

6.5 Conditional distribution: Continuous Case

- We have

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

and similarly

$$f(x, y) = f_{Y|X}(y|x)f_X(x)$$

- Hence we obtain

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

which holds if $f_Y(y) > 0$.

6.5 Conditional Expectation and Variance: Continuous Case

- We can define the mean, variance of the conditional p.d.f.
- The conditional mean is given by

$$E(X|Y = y) = \int x f_{X|Y}(x|y) dx$$

- The conditional variance is given by

$$\begin{aligned} \text{Var}(X|Y = y) &= E((X - E(X|Y = y))^2 | Y = y) \\ &= E(X^2 | Y = y) - E(X|Y = y)^2 \end{aligned}$$

where

$$E(X^2 | Y = y) = \int x^2 f_{X|Y}(x|y) dx$$

- $E(X|Y = y)$ and $\text{Var}(X|Y = y)$ are functions of x and y respectively, but $E(X|Y)$ and $\text{Var}(X|Y)$ are random variables.

6.5 Example

- Find the conditional pdf of Y given X and $E(Y|X = x)$ when their joint density is

$$f(x, y) = xe^{-x(y+1)}, \quad \text{for } x, y \geq 0$$

- We have for $0 \leq x \leq \infty$

$$f_X(x) = \int_0^{\infty} xe^{-x(y+1)} dy = e^{-x}$$

so that for $0 \leq y \leq \infty$,

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} = xe^{-xy}$$

- It follows that

$$E(Y|X = x) = \int_0^{\infty} yf_{Y|X}(y|x) dy = \frac{1}{x}.$$

6.6 Order Statistics

- Let $X_1, X_2, \dots, X_n$ be n independent and identically distributed continuous random variables having a common density f and distribution function F .
- The ordered values $X_{(1)} \leq X_{(2)} \leq \dots X_{(n)}$ are known as the order statistics.
- What's the distribution of $X_{(1)}$ and $X_{(n)}$ respectively?

6.7 Joint probability distribution of functions of random variables

- **Example:** Suppose X and Y are two common factors that drive the dynamic of the Eurodollar U and Japanese Yen V . Then both U and V are the functions of X and Y .
- **Example:** Suppose both demand U and supply V of some commodity are the functions of the commodity price X and consumer income Y .
- We are interested at the bivariate transformation:

$$U = g_1(X, Y), V = g_2(X, Y)$$

- How to find the joint distribution of (U, V) based on the distribution of (X, Y) ?

6.7 Joint probability distribution of functions of random variables

- **Definition: Jacobian matrix and Jacobian.** Consider the bivariate transformation: $U = g_1(X, Y)$, $V = g_2(X, Y)$, where g_1, g_2 are continuously differentiable with respect to x, y . Then the 2×2 matrix

$$J_{UV}(x, y) = \begin{pmatrix} \frac{\partial g_1(X, Y)}{\partial x} & \frac{\partial g_1(X, Y)}{\partial y} \\ \frac{\partial g_2(X, Y)}{\partial x} & \frac{\partial g_2(X, Y)}{\partial y} \end{pmatrix}$$

is called the *Jacobian matrix* of (U, V) . Its determinant is called the *Jacobian*.

6.7 Joint probability distribution of functions of random variables

- **Definition: Inverse Function.** Suppose A and B are subsets of $\mathbb{R}^2$. Suppose transformation $(U, V) = (g_1(X, Y), g_2(X, Y))$ is continuously differentiable with the determinant of $J_{UV}(x, y)$ not zero on A . Then the following functions h_1, h_2 exist:
 $(X, Y) = (h_1(U, V), h_2(U, V))$, where $(h_1, h_2) : B \rightarrow A$ are continuously differentiable on B such that

$$\begin{aligned}h_1(g_1(x, y), g_2(x, y)) &= x, \\h_2(g_1(x, y), g_2(x, y)) &= y.\end{aligned}$$

We call (h_1, h_2) the *inverse functions* of (g_1, g_2) .

- **Theorem:** The Jacobian of (X, Y) satisfies

$$\det[J_{XY}(u, v)] = 1 / \det[J_{UV}(h_1(u, v), h_2(u, v))].$$

6.7 Joint probability distribution of functions of random variables

- **Theorem: Bivariate Transformation.** Let (X, Y) be a bivariate continuous random variable with pdf $f_{XY}(x, y)$, and let $A = \{(x, y) \mid f_{XY}(x, y) > 0\}$. Define $U = g_1(X, Y)$, $V = g_2(X, Y)$, where $(g_1, g_2) : A \rightarrow B$ is a 1-1 and continuously differentiable function with $\det[J_{UV}(x, y)] \neq 0$. Let (h_1, h_2) be the inverse function of (g_1, g_2) , that is $X = h_1(U, V)$, $Y = h_2(U, V)$. Then

$$\begin{aligned} f_{UV}(u, v) &= \begin{cases} f_{XY}(h_1(u, v), h_2(u, v)) |\det[J_{XY}(u, v)]| & \text{for } (u, v) \in B \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{f_{XY}(h_1(u, v), h_2(u, v))}{|\det[J_{U, V}(h_1(u, v), h_2(u, v))]|} & \text{for } (u, v) \in B, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

6.7 Joint probability distribution of functions of random variables

- **Example:** Suppose X and Y are independent following $N(0, \sigma^2)$. Let $U = X + Y$, $V = X - Y$. Find the joint pdf $f_{UV}(u, v)$.
- **ANS:**
 - The joint pdf of (U, V) is

$$\begin{aligned}f_{UV}(u, v) &= f_{XY}(x, y) \times |\det[J_{XY}(u, v)]| \\&= \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \times \frac{1}{2} \\&= \frac{1}{4\pi\sigma^2} e^{-\frac{u^2+v^2}{4\sigma^2}}.\end{aligned}$$

- U and V are independent.

6.7 Joint probability distribution of functions of random variables

- **Example: Sum of Random Variables.** Assume (X, Y) have joint pdf $f_{XY}(x, y)$, find the pdf of $U = X + Y$.
- **ANS:**
 - Let $V = Y$, then $|\det[J_{UV}]| = 1$.
 - The pdf of (U, V) is

$$f_{UV}(u, v) = f_{XY}(u - v, v).$$

- Therefore, the marginal pdf of U is

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(u - v, v) dv.$$

6.7 Joint probability distribution of functions of random variables

- **Example: Sum of Two Uniform Random Variables.** Let the joint density of (X, Y) be

$$f_{XY}(x, y) = \begin{cases} 1 & 0 < x < 1, 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}.$$

Find the pdf of $U = X + Y$.

- **ANS:**

- For $0 < u \leq 1$,

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(u - v, v) dv = \int_0^u 1 dv = u.$$

- For $1 < u < 2$,

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(u - v, v) dv = \int_{u-1}^1 1 dv = 2 - u.$$

6.7 Joint probability distribution of functions of random variables

- **Example: Product of Two Random Variables.** Assume (X, Y) have joint pdf $f_{XY}(x, y)$, find the pdf of $U = XY$.

- **ANS:**

- Let $V = X$. Then $X = V$, $Y = U/V$. The Jacobian of this transformation is

$$\det[J_{XY}(u, v)] = \begin{vmatrix} 0 & 1 \\ \frac{1}{v} & -\frac{u}{v^2} \end{vmatrix} = \frac{1}{v}.$$

- The joint pdf of (U, V) is

$$f_{UV}(u, v) = \frac{1}{|v|} f_{XY}(v, u/v).$$

- The marginal pdf of U is

$$f_U(u) = \int_{-\infty}^{\infty} \frac{1}{|v|} f_{XY}(v, u/v) dv.$$

6.7 Joint probability distribution of functions of random variables

- **Example:** X and Y are independent following $N(0, \sigma^2)$. Let $U = X^2 + Y^2$, $V = X / \sqrt{X^2 + Y^2}$. (a) Find $f_{UV}(u, v)$. (b) Show that U and V are independent.
- **ANS:**
 - Let $W = Y^2$. Then

$$f_W(w) = \frac{1}{\sqrt{2\pi w}\sigma} e^{-\frac{w}{2\sigma^2}}$$

for $w \geq 0$.

- $(U, V) = (X^2 + W, X / \sqrt{X^2 + W})$ is a 1-1 mapping from $(-\infty, \infty) \times [0, \infty)$ to $[0, \infty) \times [-1, 1]$. Its inverse function is $X = \sqrt{UV}$, $W = U - UV^2$.

6.7 Joint probability distribution of functions of random variables

- The joint pdf of (U, V) is

$$\begin{aligned}f_{UV}(u, v) &= f_{XW}(x, w) \times |\det[J_{XW}(u, v)]| \\&= \frac{1}{2\pi\sigma^2\sqrt{w}} e^{-\frac{x^2+w}{2\sigma^2}} \times \sqrt{u} \\&= \frac{1}{2\pi\sigma^2\sqrt{1-v^2}} e^{-\frac{u}{2\sigma^2}}.\end{aligned}$$

- U and V are independent.

6.7 Joint probability distribution of functions of random variables

- Suppose (X, Y) has a bivariate normal PDF

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\}$$

- What's the marginal pdf $f_X(x)$, $f_Y(y)$?
- What's the conditional pdf $f_{X|Y}(x|y)$, $f_{Y|X}(y|x)$?
- If X and Y are normally distributed, are X and Y necessarily jointly normal?