Lecture 4: Multivariate Distributions

PhD Mathematics II: Probability

Juan Llavador Peralt

IIES

November 10, 2024

Outline

Bivariate Distributions

Independence

Random Vectors

Conditional Distributions

Borel Sets in $\mathbb{R}^2$ and $\mathbb{R}^n$

- We defined the Borel sets in $\mathbb{R}$, we can extend this definition to $\mathbb{R}^2$ and $\mathbb{R}^n$:

$$\mathcal{B}(\mathbb{R}^2) = \sigma\left(\{(a,b] \times (c,d] : a,b,c,d \in \mathbb{R}\}\right)$$

- Similarly:

$$\mathcal{B}(\mathbb{R}^n) = \sigma\left(\{(a_1,b_1] \times \cdots \times (a_n,b_n] : a_i, b_i \in \mathbb{R}\}\right)$$

Bivariate Distributions – The CDF

- Let X and Y be two random variables

- We write

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x \cap Y \leq y) = \mathbb{P}(X \leq x, Y \leq y)$$

Properties:

1. $\lim_{x,y\to\infty} F_{X,Y}(x, y) = 1$
   $\lim_{x\to-\infty} F_{X,Y}(x, y) = 0, \forall y$
   $\lim_{y\to-\infty} F_{X,Y}(x, y) = 0, \forall x$

2. Right continuous in x: $\lim_{h\to 0^+} F_{X,Y}(x + h, y) = F_{X,Y}(x, y), \forall y$
   Right continuous in y: $\lim_{h\to 0^+} F_{X,Y}(x, y + h) = F_{X,Y}(x, y), \forall x$

3. Monotonicity in x: $x_1 \leq x_2 \Rightarrow F_{X,Y}(x_1, y) \leq F_{X,Y}(x_2, y), \forall y$
   Monotonicity in y: $y_1 \leq y_2 \Rightarrow F_{X,Y}(x, y_1) \leq F_{X,Y}(x, y_2), \forall x$

Recovering the marginal CDFs

- Recall that when we write $\mathbb{P}(X \leq x)$ we mean $\mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\})$
- So as $x \to \infty$, we have $\mathbb{P}(X \leq \infty) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq \infty\}) = \mathbb{P}(\Omega) = 1$
- For example, the marginal CDF of X is

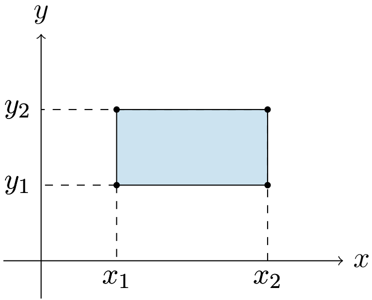
$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y)$$

- Similarly, the marginal CDF of Y is

$$F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x, y)$$

A simple example

- We are interested in the probability that X, Y take values in a (Borel) subset B of $\mathbb{R}^2$
- The simplest case is when B is a rectangle $B = (x_1, x_2] \times (y_1, y_2]$



How do we compute $\mathbb{P}(X \in (x_1, x_2], Y \in (y_1, y_2])$?

$$\mathbb{P}(X \in (x_1, x_2], Y \in (y_1, y_2]) = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_1, y_2) - F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1)$$

Bivariate PMF

- When both X and Y are **discrete**, we can define the **joint PMF** as

$$f_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)$$

- We can thus recover the CDF as

$$F_{X,Y}(x,y) = \sum_{x' \leq x, y' \leq y} f_{X,Y}(x',y')$$

- How can we recover the marginal PMFs?

$X \backslash Y$	y_1	y_2	y_3
x_1	p_{11}	p_{12}	p_{13}
x_2	p_{21}	p_{22}	p_{23}

- $f_X(x) = \sum_y f_{X,Y}(x,y)$ and $f_Y(y) = \sum_x f_{X,Y}(x,y)$

Bivariate PDF

- Suppose both X and Y are **jointly-continuous**,
- The PDF is an integrable function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ such that

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v) du dv$$

- With the following properties:
  1. $f_{X,Y}(x,y) \geq 0$ for all $(x,y) \in \mathbb{R}^2$
  2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1$
  3. For any Borel set $B \subset \mathbb{R}^2$, $\mathbb{P}((X,Y) \in B) = \int \int_B f_{X,Y}(x,y) dx dy$

Recovering the marginal PDFs

- The marginal PDFs are defined as

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

- Similarly,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx$$

- Question: Say we know $f_X(x)$ and $f_Y(y)$, can we recover $f_{X,Y}(x, y)$?
- Not in general, only if X and Y are **independent**

Joint Moments

- Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a well-behaved function and X, Y two random variables:

$$\mathbb{E}[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y) f_{X,Y}(x,y), & \text{discrete case,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\, dx dy, & \text{continuous.} \end{cases}$$

- Like in the univariate case, we will use the common notation:

$$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) dF_{X,Y}(x,y)$$

- We call the **joint moments** of X, Y the expectations of $X^m Y^n$ for $m, n \in \mathbb{N}$
- And the **joint central moments** are defined as $\mathbb{E}[(X - \mathbb{E}[X])^m (Y - \mathbb{E}[Y])^n]$
- Example: the **covariance** is $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

Covariance Properties

- **Symmetry**:

$$\mathsf{Cov}(X, Y) = \mathsf{Cov}(Y, X)$$

- **Bilinearity**:

$$\mathsf{Cov}(X_1 + X_2, Y) = \mathsf{Cov}(X_1, Y) + \mathsf{Cov}(X_2, Y)$$

and

$$\mathsf{Cov}(aX, Y) = a\mathsf{Cov}(X, Y)$$

- **Variance**:

$$\mathsf{Cov}(X, X) = \mathsf{Var}(X)$$

and

$$\mathsf{Var}[X + Y] = \mathsf{Var}(X) + \mathsf{Var}(Y) + 2\mathsf{Cov}(X, Y)$$

- **Independence**: If X and Y are independent, then

$$\mathsf{Cov}(X, Y) = 0$$

Correlation

- The **correlation** between two random variables X and Y is defined as

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

- For rvs. X, Y with finite second moments, we have $-1 \leq \text{Corr}(X, Y) \leq 1$
- With $|\text{Corr}(X, Y)| = 1$ iff $Y = aX + b$ for some $a, b \in \mathbb{R}$

Proof

- Define $Z = Y - aX$, then

$$0 \leq \mathsf{Var}(Z) = \mathsf{Var}(Y - aX) = \mathsf{Var}(Y) + a^2\mathsf{Var}(X) - 2a\mathsf{Cov}(X, Y) \\ = h(a)$$

- Since $h(a) \geq 0$ for all a, it has at most one root

- That is

$$\begin{aligned} 0 \geq & 4\mathsf{Cov}(X, Y)^2 - 4\mathsf{Var}(X)\mathsf{Var}(Y) = 4\left[\mathsf{Cov}(X, Y)^2 - \mathsf{Var}(X)\mathsf{Var}(Y)\right] \\ \iff & \mathsf{Cov}(X, Y)^2 \leq \mathsf{Var}(X)\mathsf{Var}(Y) \\ \iff & \frac{\mathsf{Cov}(X, Y)^2}{\mathsf{Var}(X)\mathsf{Var}(Y)} \leq 1 \\ \iff & |\mathsf{Corr}(X, Y)| \leq 1 \end{aligned}$$

- Finally, note that $\mathsf{Var}(Z) = 0$ when $Y = aX + b$

Multivariate Generalization (1/2)

For n random variables $X_1, \ldots, X_n$, we have analogous definitions:

1. The joint CDF is a function $F_{X_1,\ldots,X_n} : \mathbb{R}^n \to [0,1]$ such that

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n);$$

2. The marginal CDFs are, for any $j = 1, \ldots, n$, the functions

$$F_{X_j}(x_j) = F_{X_1,\ldots,X_n}(\infty, \ldots, \infty, x_j, \infty, \ldots, \infty);$$

Multivariate Generalization (2/2)

3. The marginal PMF or PDF are, for any $j = 1, \ldots, n$, the functions

$$f_{X_j}(x_j) = \begin{cases} \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n), & \text{discrete case,} \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_{j-1} dx_{j+1} \ldots dx_n, & \text{continuous;} \end{cases}$$

4. If g is a well-behaved function $g : \mathbb{R}^n \to \mathbb{R}$, then

$$\mathbb{E}[g(X_1, \ldots, X_n)] = \begin{cases} \sum_{x_1} \cdots \sum_{x_n} g(x_1, \ldots, x_n) f_{X_1,\ldots,X_n}(x_1, \ldots, x_n), & \text{discrete,} \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n, & \text{continuo} \end{cases}$$

Joint MGFs

- The **joint MGF** of two random variables X, Y is defined as

$$M_{X,Y}(t_1, t_2) = \mathbb{E}[e^{t_1 X + t_2 Y}]$$

- We can Taylor expand the exponential function to recover the moments of X, Y:

$$\begin{aligned} M_{X,Y}(t_1, t_2) &= \mathbb{E}\left[\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{t_1^m}{m!}\frac{t_2^n}{n!}X^m Y^n\right] \\ &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{t_1^m}{m!}\frac{t_2^n}{n!}\mathbb{E}[X^m Y^n] \end{aligned}$$

- Can recover the m, n-th moments of X, Y as $\frac{\partial^{m+n} M_{X,Y}(t_1,t_2)}{\partial t_1^m \partial t_2^n}|_{t_1=t_2=0}$

Independence of two random variables

- Two random variables X, Y are **independent** iff the events $\{X \leq x\}$ and $\{Y \leq y\}$ are mutually independent for all x, y:

$$F_{X,Y}(x,y) = F_X(x)F_Y(y), \forall x, y$$

- Equivalently, iff

$$f_{X,Y}(x,y) = f_X(x)f_Y(y), \forall x, y$$

- The random variables $X_1, \ldots, X_n$ are **mutually independent** iff the events $\{X_1 \leq x_1\}, \ldots, \{X_n \leq x_n\}$ are independent for all $x_1, \ldots, x_n$:

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = F_{X_1}(x_1)\ldots F_{X_n}(x_n)$$

- Equivalently, iff

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1)\ldots f_{X_n}(x_n)$$

Moments and Independence

- A motivating example:

$$
\begin{aligned}
\mathbb{E}[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_X(x) f_Y(y)\,dxdy \\
&= \left(\int_{-\infty}^{\infty} x f_X(x)\,dx\right)\left(\int_{-\infty}^{\infty} y f_Y(y)\,dy\right) = \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}
$$

- In general, if a function $g(x, y) = h(x)k(y)$ then for X, Y independent:

$$\mathbb{E}[g(X,Y)] = \mathbb{E}[h(X)]\mathbb{E}[k(Y)]$$

- If X, Y are independent, then $M_{X,Y}(t_1, t_2) = M_X(t_1)M_Y(t_2)$

Sum of two random variables

- Let X and Y be two random variables with joint density $f_{X,Y}(x,y)$
- Let $Z = X + Y$
- Then:

$$f_Z(z) = \begin{cases} \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)dx, & \text{continuous case,} \\ \sum_x f_{X,Y}(x, z-x), & \text{discrete case.} \end{cases}$$

- **Proof:** (discrete case): Define $Z = X + Y$:

$$\{Z = z\} = \{X + Y = z\} = \bigcup_u \{X = u, Y = z - u\}$$

$$f_Z(z) = \mathbb{P}(Z = z) = \sum_u \mathbb{P}(X = u, Y = z - u)$$

Convolution

- If X and Y are **independent**, then $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ and

$$f_Z(z) = \begin{cases} \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx, & \text{continuous case,} \\ \sum_x f_X(x)f_Y(z-x), & \text{discrete case.} \end{cases}$$

- This operation is called **convolution** and we denote it as $f_Z = f_X * f_Y$
- In the case of n independent random variables and $S = X_1 + \cdots + X_n$, we have

$$f_S = f_{X_1} * \cdots * f_{X_n}$$

Example: Sum of two Exponentials

- Let $X \sim \text{Exp}(\lambda)$ and $Y \sim \text{Exp}(\mu)$ be two independent exponential random variables
- The PDF of X is $f_X(x) = \lambda e^{-\lambda x}$ and the PDF of Y is $f_Y(y) = \mu e^{-\mu y}$
- The PDF of $Z = X + Y$ is

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx \\
&= \int_0^z \lambda e^{-\lambda x} \mu e^{-\mu(z-x)} dx \\
&= \lambda \mu e^{-\mu z} \int_0^z e^{-(\lambda+\mu)x} dx \\
&= \frac{\lambda \mu}{\lambda + \mu} e^{-\mu z} \left(1 - e^{-(\lambda+\mu)z}\right)
\end{aligned}$$

Random Vector – Definition and Notation

Random vector: An n-dimensional vector of random variables, i.e., a function

$$\boldsymbol{X} = (X_1, \ldots, X_n)^T : \Omega \to \mathbb{R}^n.$$

The CDF, PMF or PDF, and MGF of a random vector are the joint CDF, PMF or PDF, and MGF of $X_1, \ldots, X_n$, so for any $\boldsymbol{x} = (x_1, \ldots, x_n), \boldsymbol{t} = (t_1, \ldots, t_n) \in \mathbb{R}^n$:

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{X_1,\ldots,X_n}(x_1, \ldots, x_n),$$
$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n),$$
$$M_{\boldsymbol{X}}(\boldsymbol{t}) = M_{X_1,\ldots,X_n}(t_1, \ldots, t_n).$$

Expectation of a random vector: The expectation of a random vector is a vector of the expectations, i.e., it is taken element by element:

$$\mathbb{E}[\boldsymbol{X}] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{pmatrix}.$$

Random Matrix

- Similarly, a **random matrix** is a matrix whose entries are random variables:

$$\boldsymbol{W} = \begin{pmatrix} W_{11} & \dots & W_{1n} \\ \vdots & \ddots & \vdots \\ W_{m1} & \dots & W_{mn} \end{pmatrix}$$

- With joint CDF, PMF or PDF, and MGF defined as the joint CDF, PMF or PDF, and MGF of the entries

- The expectation of a random matrix is a matrix of the expectations of the entries

Variance of a Random Vector

- The **variance-covariance matrix** of an $n \times 1$ random vector $\boldsymbol{X}$ is the $n \times n$ matrix:

$$\mathsf{Var}[\boldsymbol{X}] = \mathbb{E}[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^T]$$

$$\mathsf{Var}[\boldsymbol{X}] = \begin{pmatrix} \mathsf{Var}[X_1] & \mathsf{Cov}[X_1, X_2] & \dots & \mathsf{Cov}[X_1, X_n] \\ \mathsf{Cov}[X_2, X_1] & \mathsf{Var}[X_2] & \dots & \mathsf{Cov}[X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{Cov}[X_n, X_1] & \mathsf{Cov}[X_n, X_2] & \dots & \mathsf{Var}[X_n] \end{pmatrix}$$

- **Question:** What $\mathsf{Var}[\boldsymbol{X}]$ if $\boldsymbol{X}$ is independent and identically distributed (i.i.d.)? $\sigma^2 I_n$

The Variance-Covariance Matrix is PSD

- The variance-covariance matrix is **positive semi-definite** (PSD)
- For any vector $\boldsymbol{a} \in \mathbb{R}^n$, we have: $\boldsymbol{a}^T \text{Var}[\boldsymbol{X}]\boldsymbol{a} \geq 0$
- **Proof:** Let $\boldsymbol{X}$ be a random vector and $\boldsymbol{a} \in \mathbb{R}^n$. Then

$$\begin{aligned}
\text{Var}[\boldsymbol{a}^T \boldsymbol{X}] &= \mathbb{E}\left[(\boldsymbol{a}^T \boldsymbol{X} - \mathbb{E}[\boldsymbol{a}^T \boldsymbol{X}])(\boldsymbol{a}^T \boldsymbol{X} - \mathbb{E}[\boldsymbol{a}^T \boldsymbol{X}])^T\right] \\
&= \mathbb{E}\left[(\boldsymbol{a}^T (\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}]))(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^T \boldsymbol{a}\right] \\
&= \boldsymbol{a}^T \text{Var}[\boldsymbol{X}]\boldsymbol{a} \geq 0
\end{aligned}$$

Motivation for Conditional Distributions

- Suppose we have two random varianles X and Y and we know their joint distribution $f_{X,Y}(x,y)$
- Suppose we know that $X = x$, what can we say about Y?
- We now how to compute conditional probabilities:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

- Intuitively, for discrete RVs we can use $\{X = x\}$ and $\{Y = y\}$ and apply the same formula

Conditional PMF

- Let X and Y be two **discrete** random variables with joint PMF $f_{X,Y}(x,y)$
- The **conditional PMF** of Y given $X = x$ is defined as

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

- Is this a valid PMF?

$$\sum_y f_{Y|X}(y|x) = \sum_y \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{1}{f_X(x)} \sum_y f_{X,Y}(x,y) = 1$$

- We can only condition on $X = x$ if $f_X(x) > 0$, that is if x is in the support of X

Conditional PDF

- Let X and Y be two **continuous** random variables with joint PDF $f_{X,Y}(x,y)$
- The **conditional PDF** of Y given $X = x$ is defined as

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

- Is this a valid PDF?

$$\int_{-\infty}^{\infty} f_{Y|X}(y|x)dy = \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_X(x)}dy = \frac{1}{f_X(x)} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = 1$$

- We can only condition on $X = x$ if $f_X(x) > 0$, that is if x is in the support of X

Conditional CDF

- The **conditional CDF** of Y given $X = x$ is defined as

$$F_{Y|X}(y|x) = \mathbb{P}(Y \leq y|X = x) = \begin{cases} \sum_{y' \leq y} f_{Y|X}(y'|x), & \text{discrete case,} \\ \int_{-\infty}^{y} f_{Y|X}(y'|x)dy', & \text{continuous.} \end{cases}$$

Example: Hurricanes reaching land

- We are interested in modelling the number of hurricanes reaching land in a given year Y
- Suppose that we know that number of hurricanes $N \sim$ Poisson(λ): $f_N(n) = \frac{e^{-\lambda}\lambda^n}{n!}$
- And that each hurricane has a probability p of reaching land $Y|N = n \sim$ Binomial(n, p)
- One can show that $Y \sim$ Poisson(λp)
- You can prove this directly using the definition of conditional PMF
- We will instead use this to illustrate conditional expectations

Conditional Expectation

- We have seen that $Y|X = x$ is a perfectly valid random variable
- We can thus define the **conditional expectation** of Y **given** $X = x$ as

$$\mathbb{E}[Y|X = x] = \begin{cases} \sum_y y f_{Y|X}(y|x), & \text{discrete case,} \\ \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy, & \text{continuous.} \end{cases}$$

- Define the function $g(x) = \mathbb{E}[Y|X = x]$, then the **conditional expectation** of Y given X is

$$\mathbb{E}[Y|X]$$

- That is, $\mathbb{E}[Y|X]$ is a **random variable** that depends on X

Law of Iterated Expectations

- The **law of iterated expectations** states that for any two random variables X and Y:

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$$

- **Proof:**

$$\begin{aligned}
\int_{-\infty}^{\infty} \mathbb{E}[Y|X = x] dF_X(x) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} y dF_{Y|X}(y|x) \right) dF_X(x) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y dF_{X,Y}(x, y) \\
&= \int_{-\infty}^{\infty} y dF_Y(y) = \mathbb{E}[Y]
\end{aligned}$$

- Also works for a function of Y: $\mathbb{E}[g(Y)] = \mathbb{E}[\mathbb{E}[g(Y)|X]]$

Example: Hurricanes reaching land

- We have seen that $Y \sim \text{Poisson}(\lambda p)$
- We can compute the conditional expectation of Y given $N = n$ as

$$\mathbb{E}[Y|N = n] = np$$

- And the law of iterated expectations gives us

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|N]] = \mathbb{E}[Np] = \lambda p$$

LOTUS for Conditional Expectations

- $g : \mathbb{R}^2 \to \mathbb{R}$ is a well-behaved real function
- Then, the **conditional expectation** of $g(Y, X)$ given $X = x$ is

$$\mathbb{E}[g(Y,X)|X=x] = \int_{-\infty}^{\infty} g(y,x) dF_{Y|X}(y|x)$$

- And the **conditional expectation** of $g(Y, X)$ given X is

$$\mathbb{E}[g(Y,X)|X]$$

- Example (Taking out what is known): $\mathbb{E}[XY|X] = X\mathbb{E}[Y|X]$

Conditional Moments

- The **r-th conditional moments** of Y given $X = x$ are defined as

$$\mathbb{E}[Y^r|X = x] = \int_{-\infty}^{\infty} y^r dF_{Y|X}(y|x)$$

- And the **r-th conditional central moments** of Y given $X = x$ are defined as

$$\mathbb{E}[(Y - \mathbb{E}[Y|X = x])^r|X = x] = \int_{-\infty}^{\infty} (y - \mathbb{E}[Y|X = x])^r dF_{Y|X}(y|x)$$

- Similarly as before, we can define the **r-th conditional moments** of Y given X as

$$\mathbb{E}[Y^r|X]$$

  and the **r-th conditional central moments** of Y given X as

$$\mathbb{E}[(Y - \mathbb{E}[Y|X])^r|X]$$

Law of Iterated Variances

- We showed that we can recover the expectation of Y as $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$
- For the variance, we have a similar result:

$$\text{Var}[Y] = \mathbb{E}[\text{Var}[Y|X]] + \text{Var}[\mathbb{E}[Y|X]]$$

- **Proof:**

$$\begin{aligned}
\text{Var}[Y] &= \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \\
&= \mathbb{E}[\mathbb{E}[Y^2|X]] - \mathbb{E}[\mathbb{E}[Y|X]]^2 \\
&= \mathbb{E}[\text{Var}[Y|X]] + \mathbb{E}[\mathbb{E}[Y|X]^2] - \mathbb{E}[\mathbb{E}[Y|X]]^2 \\
&= \mathbb{E}[\text{Var}[Y|X]] + \text{Var}[\mathbb{E}[Y|X]]
\end{aligned}$$

- Hurricane, $X \sim \text{Poisson}(\lambda)$
  - $Y|X = x \sim \text{Binomial}(x, p)$, $X \sim \text{Poisson}(\lambda)$
  - $\mathbb{E}[Y|X = x] = xp$, $\text{Var}[Y|X = x] = xp(1-p)$
  - $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Xp] = \lambda p$
  - $\text{Var}[Y] = \mathbb{E}[\text{Var}[Y|X]] + \text{Var}[\mathbb{E}[Y|X]] = \mathbb{E}[Xp(1-p)] + \text{Var}[\lambda p] = \lambda p(1-p) + \lambda p = \lambda p$

Conditional MGFs

- The **conditional MGF** of Y given $X = x$ is defined as

$$M_{Y|X}(t|x) = \mathbb{E}[e^{tY}|X = x]$$

- We can recover the moments of Y given $X = x$ as

$$\frac{\partial^n M_{Y|X}(t|x)}{\partial t^n}|_{t=0} = \mathbb{E}[Y^n|X = x]$$

- Note that

$$M_Y(t) = \mathbb{E}[M_{Y|X}(t|X)]$$

Conditional MGFs in the Hurricane Example

- $M_{Y|N}(t|N=n) = (1-p+pe^t)^n$, $M_X(t) = e^{\lambda(e^t-1)}$
- Then $M_Y(t) = \mathbb{E}[M_{Y|N}(t|N)] = \mathbb{E}[(1-p+pe^t)^N] = \mathbb{E}[N\log(1-p+pe^t)]$
- We can compute this using the MGF of N: $M_Y(t) = e^{\lambda p(e^t-1)}$
- This is the MGF of a Poisson distribution with parameter λp

Conditioning for random vectors

- Let $\boldsymbol{X}, \boldsymbol{Y}$ be two random vectors
- We can define the **conditional distribution** of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$ as

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

- And the expectation of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$ as

$$\mathbb{E}[\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x}] = \begin{pmatrix} \mathbb{E}[Y_1|\boldsymbol{X} = \boldsymbol{x}] \\ \vdots \\ \mathbb{E}[Y_n|\boldsymbol{X} = \boldsymbol{x}] \end{pmatrix}$$

- The **conditional expectation** of $\boldsymbol{Y}$ given $\boldsymbol{X}$ is

$$\mathbb{E}[\boldsymbol{Y}|\boldsymbol{X}] = \begin{pmatrix} \mathbb{E}[Y_1|\boldsymbol{X}] \\ \vdots \\ \mathbb{E}[Y_n|\boldsymbol{X}] \end{pmatrix}$$

Transformation of Continuous Random Variables (1/2)

We are interested in transforming one pair of random variables into another pair of random variables.

- Consider pairs of random variables (U, V) and (X, Y).
- Suppose that X and Y are both functions of U and V:

$$X = g_1(U, V), \quad Y = g_2(U, V).$$

- Suppose g is well-behaved and invertible.
- We use the inverse transformation:

$$U = h_1(X, Y), \quad V = h_2(X, Y).$$

- The overall transformation is g, so $(X, Y) = g(U, V)$, and the inverse is h, so $(U, V) = g^{-1}(X, Y) = h(X, Y)$.
- Then, if (U, V) are continuous random variables with support D, and $(X, Y) = g(U, V)$, the joint density of X and Y is

$$f_{X,Y}(x, y) = \begin{cases} f_{U,V}(h(x, y))\,|J_h(x, y)| & \text{for } (x, y) \in R, \\ 0 & \text{otherwise.} \end{cases}$$

Transformation of Continuous Random Variables (2/2)

This is referred to as the change-of-variables formula.

- The Jacobian of the inverse transformation, $J_h(x, y)$, is given by:

$$J_h(x,y) = \begin{vmatrix} \frac{\partial}{\partial x} h_1(x,y) & \frac{\partial}{\partial x} h_2(x,y) \\ \frac{\partial}{\partial y} h_1(x,y) & \frac{\partial}{\partial y} h_2(x,y) \end{vmatrix}$$

- This simplifies to:

$$J_h(x,y) = \frac{\partial}{\partial x} h_1(x,y) \frac{\partial}{\partial y} h_2(x,y) - \frac{\partial}{\partial x} h_2(x,y) \frac{\partial}{\partial y} h_1(x,y).$$

- The Jacobian can be expressed in terms of the Jacobian of the original transformation J_g:

$$J_g(u,v) = \begin{vmatrix} \frac{\partial}{\partial u} g_1(u,v) & \frac{\partial}{\partial u} g_2(u,v) \\ \frac{\partial}{\partial v} g_1(u,v) & \frac{\partial}{\partial v} g_2(u,v) \end{vmatrix}.$$

Back

Proof in the Continuous Case

- Let $Z = X + Y$, $U = X$ or $X = U$, $Y = Z - U$
- hence $\frac{\partial X}{\partial U} = 1$, $\frac{\partial X}{\partial Z} = 0$, $\frac{\partial Y}{\partial U} = -1$, $\frac{\partial Y}{\partial Z} = 1$
- The Jacobian of the inverse transformation is:

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

- The joint density of U, Z is

$$f_{U,Z}(u, z) = f_{X,Y}(u, z - u) \times 1 = f_{X,Y}(u, z - u)$$

- Then, the density of $Z = X + Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(u, z - u) du$$