

Week 8 and 9 Recap

Pranav Rao

November 11, 2023

1 Joint Distributions

- Given two random variables defined on the same probability space, the **joint probability distribution** is the corresponding probability distribution on all possible pairs of outputs.
- A joint probability distribution “encodes” two other types of probability distributions:
 - **Marginal distributions:** the distributions of each individual random variable.
 - **Conditional distributions:** deal with how the outputs of one random variable are distributed when given information about the other random variable.
- We use the **covariance** to measure the relationship between two random variables (definition depends on types of random variables, see below).
 - However, the covariance is in the same units as the variances, and therefore is affected by the units that we choose to measure the data.
 - So, we also define the **correlation**, which is unitless (definition depends on types of random variables, see below)
 - * The correlation has values on $[-1, 1]$:
 - A correlation of 0 indicates there is no linear relationship between the two random variables (i.e. they are *linearly independent*).
 - A correlation of ± 1 indicates a perfect linear positive or negative relationship.
 - Correlation values near ± 1 imply a “strong” correlation, values near 0 imply a “weak” correlation, and values in between imply “moderate” correlation.
- Like regular random variables/distributions, there are two types of joint distributions, which we will look into in separate sections of this recap:

- Discrete
- Continuous

1.1 Joint Distributions of Discrete Random Variables

- Consider two discrete random variables X and Y .
- The **joint probability mass function**, defined as $P_{XY}(X = x, Y = y)$ is normally given in a table, and must satisfy two properties for every x, y :

- $0 \leq P_{XY}(x, y) \leq 1$
- $\sum_x \sum_y P_{XY}(x, y) = 1$

- The **marginal probability mass functions** $P_X(X = x), P_Y(Y = y)$ can be defined as such:

- $P_X(X = x) = \sum_y P_{XY}(x, y)$
- $P_Y(Y = y) = \sum_x P_{XY}(x, y)$

- The **joint cumulative distribution function** $F_{XY}(x, y)$ of two discrete random variables X, Y with PMF $P_{XY}(x, y)$ can be defined as such:

$$F_{XY}(x, y) = P_{XY}(X \leq x, Y \leq y)$$

- The **marginal cumulative distribution functions** $F_X(x)$ and $F_Y(y)$ can be defined as such:

- $F_X(x) = \lim_{y \rightarrow \infty} F_{XY}(x, y) = F_{XY}(x, \infty)$ (shorthand) for any x
- $F_Y(y) = \lim_{x \rightarrow \infty} F_{XY}(x, y) = F_{XY}(\infty, y)$ (shorthand) for any y

- **Conditional probability mass functions** require more detail and are discussed in a later section.

- The **expected value** can be calculated as $E[g(X, Y)] = \sum_x \sum_y g(x, y) \cdot F_{XY}(x, y)$,

where $g(X, Y)$ is a function of X, Y and $F_{XY}(x, y)$ is a joint PMF of X and Y . Some consequences of this are:

- When $g(X, Y) = X$, $E(g(X, Y)) = E(X) = \sum_x x P_X(x)$, where $P_X(x)$ is the marginal pmf of X

- When $g(X, Y) = Y$, $E(g(X, Y)) = E(Y) = \sum_y yP_Y(y)$, where $P_Y(y)$ is the marginal pmf of Y
- The **variance** is still defined as $Var(g(X, Y)) = E[g(X, Y)^2] - (E[g(X, Y)])^2$, where $g(X, Y)$ is a function of X, Y . Some consequences of this are:
 - $Var(X) = \sum_x x^2 P_X(x) - (E(X))^2$, where $P_X(x)$ is the marginal pmf of X
 - $Var(Y) = \sum_y y^2 P_Y(y) - (E(Y))^2$, where $P_Y(y)$ is the marginal pmf of Y
- As usual, the **standard deviation** σ_X, σ_Y , can be calculated by taking the square roots of the respective variances.
- The **covariance** $Cov(X, Y)$ of X and Y in a joint discrete distribution can be calculated as such:

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$
- The **correlation** ρ_{XY} of X and Y in a joint discrete distribution can be calculated as such:

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$
- Discrete random variables are **independent** if their joint PMF factors into a product of the marginal PMFs.

1.2 Joint Distributions of Continuous Random Variables

- Consider two continuous random variables X and Y .
- The **joint probability distribution function**, defined as a piecewise continuous function $f_{XY}(x, y)$ or sometimes just $f(x, y)$, must satisfy two properties for every x, y in the domain of f_{XY} :
 - $f_{XY}(x, y) \geq 0$
 - $\int_y \int_x f_{XY}(x, y) dx dy = 1$ (note that \int_x means to integrate over all x from $(-\infty, \infty)$), same for every other reference in this document
- The **marginal probability distribution functions** $f_X(x), f_Y(y)$ can be defined as such:
 - $f_X(x) = \int_y f_{XY}(x, y) dy$
 - $f_Y(y) = \int_x f_{XY}(x, y) dx$

- The **joint cumulative distribution function** $F_{XY}(a, b)$ of two continuous random variables X, Y with PDF $f_{XY}(x, y)$ can be defined as such:

$$F_{XY}(a, b) = f_{XY}(X \leq a, Y \leq b) = \int_{-\infty}^b \int_{-\infty}^a f_{XY}(x, y) dx dy$$

- It must satisfy some properties:
 - * $F_{XY}(\infty, \infty) = 1$
 - * $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0$
- The **marginal cumulative distribution functions** $F_X(x)$ and $F_Y(y)$ can be defined as such:
 - $F_X(x) = \lim_{y \rightarrow \infty} F_{XY}(x, y) = F_{XY}(x, \infty)$ (shorthand) for any x
 - $F_Y(y) = \lim_{x \rightarrow \infty} F_{XY}(x, y) = F_{XY}(\infty, y)$ (shorthand) for any y
- **Conditional probability distribution functions** require more detail and are discussed in a later section.
- The **expected value** can be calculated as $E[g(X, Y)] = \int_y \int_x g(x, y) f_{XY}(x, y) dx dy$, where $g(X, Y)$ is a function of X, Y and $f_{XY}(x, y)$ is a joint PDF of X and Y . Some consequences of this are:
 - When $g(X, Y) = X$, $E(g(X, Y)) = E(X) = \int_x x \cdot f_X(x) dx$, where $f_X(x)$ is the marginal pdf of X
 - When $g(X, Y) = Y$, $E(g(X, Y)) = E(Y) = \int_y y \cdot f_Y(y) dy$, where $f_Y(y)$ is the marginal pdf of Y
- The **variance** is still defined as $Var(g(X, Y)) = E[g(X, Y)^2] - (E[g(X, Y)])^2$, where $g(X, Y)$ is a function of X, Y . Some consequences of this are:
 - $Var(X) = \int_x x^2 f_X(x) dx - (E(X))^2$, where $f_X(x)$ is the marginal pdf of X
 - $Var(Y) = \int_y y^2 f_Y(y) dy - (E(Y))^2$, where $f_Y(y)$ is the marginal pdf of Y
- As usual, the **standard deviation** σ_X, σ_Y , can be calculated by taking the square roots of the respective variances.
- The **covariance** $Cov(X, Y)$ of X and Y in a joint continuous distribution can be calculated as such:

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$
- The **correlation** ρ_{XY} of X and Y in a joint continuous distribution can be calculated as such:

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$
- Continuous random variables are **independent** if their joint PDF factors into a product of the marginal PDFs.

2 Conditional Probabilities of Joint Distributions

2.1 Discrete Distributions

Note: for this section, whenever it says “given $Y = y$ ” (for any given random variable), you can also replace that with “given an event A has occurred.”

- The **conditional probability mass functions** $P_{X|Y}(x|y)$ and $P_{Y|X}(y|x)$ of a discrete distribution can be defined as such (where $P(x, y)$ is a joint PMF for X and Y):
 - “Conditional probability mass function of X given that $Y = y$ ”: $P_{X|Y}(x|y) = \frac{P(\{X=x\} \cap \{Y=y\})}{P(Y=y)} = \frac{P(x, y)}{P_Y(y)}$, provided $P_Y(y) > 0$ (or else it doesn’t exist).
 - “Conditional probability mass function of Y given that $X = x$ ”: $P_{Y|X}(y|x) = \frac{P(\{Y=y\} \cap \{X=x\})}{P(X=x)} = \frac{P(x, y)}{P_X(x)}$, provided $P_X(x) > 0$ (or else it doesn’t exist).
- Below are some properties of conditional PMFs:
 1. Conditional PMFs are valid PMFs, meaning that they satisfy the properties of a PMF (see definition of PMF for Joint Distribution above).
 2. *In general*, the conditional PMF of X given Y does not equal the conditional distribution of Y given X , i.e.:

$$P_{X|Y}(x|y) \neq P_{Y|X}(y|x)$$

3. If X and Y are independent, then:
 - $P_{X|Y}(x|y) = P_X(x)$
 - $P_{Y|X}(y|x) = P_Y(y)$
- The **conditional cumulative distribution functions** $F_{X|Y=y}(x)$ and $F_{Y|X=x}(y)$ of a discrete distribution (where $P_{X|Y}(x|y)$ is a conditional PMF for X and Y and opposite) can be written as such:
 - “Conditional CDF of X given that $Y = y$ ”: $F_{X|Y=y}(x) = P(X \leq x|Y = y) = \sum_{a \leq x} P_{X|Y}(a|Y = y)$
 - “Conditional CDF of Y given that $X = x$ ”: $F_{Y|X=x}(y) = P(Y \leq y|X = x) = \sum_{a \leq y} P_{Y|X}(a|X = x)$
 - The **conditional expected values** $E[X|Y = y]$ and $E[Y|X = x]$ can be calculated as such:

- “Conditional expected value of X , given Y ”: $\mu_{X|Y=y} = E[X|Y = y] = \sum_x x P_{X|Y}(x|y)$
- “Conditional expected value of Y , given X ”: $\mu_{Y|X=x} = E[Y|X = x] = \sum_y y P_{Y|X}(y|x)$
- You can also calculate the **conditional variances**. For example:
 - “Conditional variance of X given $Y = y$ ”: $\sigma_{X|Y=y}^2 = \text{Var}(X|Y = y) = E[X^2|Y = y] - (E[X|Y = y])^2$
 - “Conditional variance of Y given $X = x$ ”: $\sigma_{Y|X=x}^2 = \text{Var}(Y|X = x) = E[Y^2|X = x] - (E[Y|X = x])^2$

2.2 Continuous Distributions

Note: for this section, whenever it says “given $Y = y$ ” (for any given random variable), you can also replace that with “given an event A has occurred.”

- The **conditional probability distribution functions** $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$ of a continuous distribution can be defined as such (where $f(x, y)$ is a joint PDF for X and Y):
 - “Conditional probability distribution function of X given that $Y = y$ ”: $f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$.
 - “Conditional probability distribution function of Y given that $X = x$ ”: $f_{Y|X}(y|x) = \frac{f(y, x)}{f_X(x)}$.
- Below are some properties of conditional PDFs:
 1. Conditional PDFs are valid PDFs, meaning that they satisfy the properties of a PDF (see definition of PDF for Joint Distribution above).
 2. *In general*, the conditional PDF of X given Y does not equal the conditional distribution of Y given X , i.e.:

$$f_{X|Y}(x|y) \neq f_{Y|X}(y|x)$$
 3. If X and Y are independent, then:
 - $f_{X|Y}(x|y) = f_X(x)$
 - $f_{Y|X}(y|x) = f_Y(y)$
- The **conditional cumulative distribution functions** $F_{X|Y=y}(x)$ and $F_{Y|X=x}(y)$ of a discrete distribution (where $f_{X|Y}(x|y)$ is a conditional PDF for X and Y and opposite) can be written as such:

- “Conditional CDF of X given that $Y = y$ ”: $F_{X|Y=y}(x) = P(X \leq x|Y = y) = \int_{-\infty}^x f_{X|Y}(x|y)dx$
- “Conditional CDF of Y given that $X = x$ ”: $F_{Y|X=x}(y) = P(Y \leq y|X = x) = \int_{-\infty}^y f_{Y|X}(y|x)dy$
- The **conditional expected values** $E[X|Y = y]$ and $E[Y|X = x]$ can be calculated as such:
 - “Conditional expected value of X , given Y ”: $\mu_{X|Y=y} = E[X|Y = y] = \int_x x \cdot f_{X|Y}(x, y)dx$
 - “Conditional expected value of Y , given X ”: $\mu_{Y|X=x} = E[Y|X = x] = \int_y y \cdot f_{Y|X}(y, x)dy$
- You can also calculate the **conditional variances**. For example:
 - “Conditional variance of X given $Y = y$ ”: $\sigma_{X|Y=y}^2 = \text{Var}(X|Y = y) = E[X^2|Y = y] - (E[X|Y = y])^2$
 - “Conditional variance of Y given $X = x$ ”: $\sigma_{Y|X=x}^2 = \text{Var}(Y|X = x) = E[Y^2|X = x] - (E[Y|X = x])^2$

3 Finding Joint PMFs, PDFs, and CDFs of Independent Random Variables

- A **property of independent random variables**: if X and Y are independent random variables, the joint PDF, PDF, CDF, Expected Values, Variance, etc. of a function of X, Y can be obtained by *applying the same function* to the original functions’ PMFs, PDFs, CDF, etc.
- For example, assume X and Y are independent random variables. Then:
 - If $P_X(x), P_Y(y)$ are PMFs for X, Y , respectively, then the joint PMF for XY is $P_{XY}(x, y) = P_X(x) \cdot P_Y(y)$.
 - If $f_X(x), f_Y(y)$ are PDFs for X, Y , respectively, then the joint PDF for XY is $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$.
 - Same for CDF, expected value, variance, etc.

4 Two Useful Laws of Joint Distributions

4.1 Law of Total Expectation (AKA Law of Iterated Expectations)

Let X, Y be random variables in the same probability space. Then, the following holds:

$$E(X) = E_Y[E_{X|Y}(X|Y)]$$

4.2 Law of Total Variance (AKA Eve's Law)

Let X, Y be random variables in the same probability space and assume the variance of Y is finite (i.e. it is bound by a finite number). Then, the following holds:

$$Var(Y) = E_X[Var_{Y|X}(Y|X)] + Var_X(E_{Y|X}[Y|X])$$

5 Common Multivariate (Joint) Probability Distributions

5.1 Multinomial Probability Distribution

- Useful in a case where there are n independent and identically distributed trials where each trial can result in one of r possible outcomes with probabilities p_1, p_2, p_3, p_r (these probabilities must add to 1).
- The joint pmf for $X_i, i = 1, 2, \dots, r$, the frequencies (counts) of each of the r outcomes in n trials is:

$$p_{X_1, X_2, X_3 \dots X_r}(x_1, x_2, x_3 \dots x_r) = \frac{n!}{x_1! x_2! \dots x_r!} p_1^{x_1} + p_2^{x_2} + \dots + p_r^{x_r}$$

where

$$x_i \in \{0, 1, \dots, n\} (i = 1, 2, \dots, r), x_1 + x_2 + \dots + x_r = n$$

- Note: the r possible outcomes do not need to be ordered categories, but the order of the x values must be consistent with the order of the probabilities (i.e. the index i represents a specific category).
- The marginal distributions for the frequency of each category are binomial, i.e. $X_i \sim \text{Bin}(n, p)$ for $i = 1, \dots, r$.
- Associated R functions: `dmultinom`, `rmultinom`

5.2 Multivariate Normal Distribution

- Useful joint distribution of normal variables with joint pdf:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right)\right)$$

for

$$-\infty \leq x \leq +\infty, -\infty \leq y \leq +\infty$$

- Five parameters: $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho$, where ρ is the covariance between X and Y .
- Marginal PDFs of X and Y are:

$$- N(\mu_X, \sigma_X^2), N(\mu_Y, \sigma_Y^2)$$

- Conditional distributions are univariate normal:

$$- E(X|Y = y) = \mu_X + \rho\sigma_X\left(\frac{y-\mu_Y}{\sigma_Y}\right) \text{ and vice versa}$$

$$- Var(X|Y = y) = \sigma_X^2(1 - \rho^2) \text{ and vice versa}$$

- Associated R functions (part of `mvtnorm`, not standard library): `dmvnorm`, `pmvnorm`, `rmvnorm`