# Econ 210 - Probability Review

Sidharth Sah[1]

September 8, 2023

[1] Thanks to Azeem Shaikh and Max Tabord-Meehan for useful material

# Random Variables

- Random Variables are used to represent things that are uncertain - outcome of a coin flip, score of the next Superbowl, etc
- The Distribution of a rv characterizes the probability with which that rv takes on different values

# Discrete Random Variables

- Discrete random variables take on (for our purposes) a finite number of values
- Distributions of discrete rv's are called Probability Mass Functions
- Ex. If $X$ is a Bernoulli rv, it will have a pmf

$$P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

  for some $p \in [0, 1]$
- The sum of values of the pmf across all (distinct) outcomes will always be 1

# Continuous Random Variables

- Continuous random variables take on (for our purposes) a continuum of values
- Distributions of continuous rv's are called Probability Density Functions
- Ex. If $X$ is a uniform[a,b] rv, it will have a pdf

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{if otherwise} \end{cases}$$

- The integral of the pdf over the support of the rv will always equal 1

# Cumulative Distribution Functions

- All rv's have a Cumulative Distribution Function:

$$F(x) = P(X \leq x)$$

- Ex. Bernoulli cdf:

$$F(x) = \begin{cases} 0 & \textit{if } x < 0 \\ 1-p & \textit{if } 0 \leq x < 1 \\ 1 & \textit{if } x \geq 1 \end{cases}$$

- Ex. Uniform[a,b] cdf:

$$F(x) = \begin{cases} 0 & \textit{if } x < a \\ \frac{x-a}{b-a} & \textit{if } a \leq x < b \\ 1 & \textit{if } x \geq b \end{cases}$$

# Expectations

- The Expectation of a rv tells us the average value we'd get if we drew the rv from its distribution many times
- For a discrete rv that takes on $k$ values:

$$E[X] = \sum_{i=1}^{k} x_i p(x_i)$$

- For a continuous rv:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

# Expectations

- Expectations of functions of random variables work the same way
- For a discrete rv that takes on $k$ values:

$$E[g(X)] = \sum_{i=1}^{k} g(x_i)p(x_i)$$

- For a continuous rv:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

# Expectations and Probabilities

- It is often useful to go back and forth between probabilities and expectations. This is fairly simple to do:

$$P(X \in A) = E[\mathbb{1}\{X \in A\}]$$

where $A$ is some set of values and $\mathbb{1}\{\cdot\}$ is the indicator function - a function that takes on a value of 1 if its argument is true and takes on a value of 0 otherwise

# Properties of Expectations

- If $X$, $Y$ are random variables and $a$, $b$ are scalars, then
  (i) $E[a + bX] = a + bE[X]$
  (ii) $E[X + Y] = E[X] + E[Y]$
  (iii) if $X \leq Y$ (always), then $E[X] \leq E[Y]$.
- (i) Further implies that $E[a] = a$
- (iii) Further implies that if $Y \geq 0$ then $E[Y] \geq 0$

# Variance

- The Variance of a random variable is a measure of how disperse the distribution is:

$$\sigma_X^2 = Var(X) = E[(X - E[X])^2]$$

- The units of the variance are the units of X squared - slightly awkward to interpret
- The Standard Deviation, the root of the variance, has the same units as $X$, which is easier to think about:

$$\sigma_X = Std\ Dev(X) = \sqrt{Var(X)}$$

# Properties of Variance

- Alternative form of the variance:

$$\begin{aligned} E[(X-E[X])^2] &= E[X^2-2XE[X]+E[X]^2] \\ &= E[X^2]-2E[X]E[X]+E[X]^2 \\ \Rightarrow Var(X) &= E[X^2]-E[X]^2 \end{aligned}$$

- For $a$, $b$ scalars:

$$Var(a+bX) = b^2 Var(X)$$

- The latter property further implies $Var(a) = 0$

# Random Vectors/Joint Distributions

- We'll often care about how features of how two or more variables are simultaneously distributed - the Joint Distribution
- For discrete variables there will be a joint pmf:

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

and for continuous variables, there will be a joint pdf, $f(x, y)$

# Joint Distributions to Marginal Distributions

- The probability of one jointly distributed variable taking on a given value is the sum over the probabilities of all random vectors in which that variable takes on the given value. For 2 variables, if $X$ takes on $k$ values:

$$P(Y = y) = \sum_{i=1}^{k} P(Y = y, X = x_i)$$

# Covariance

- The Covariance captures whether two variables "move together" or not - if one is above average, will the other tend to also be above average?

$$\begin{aligned} Cov(X,Y) &= E[(X-E[X])(Y-E[Y])] \\ &= E[XY - XE[Y] - YE[X] - E[X]E[Y]] \\ &= E[XY] - 2E[X]E[Y] + E[X]E[Y] \\ \Rightarrow Cov(X,Y) &= E[XY] - E[X]E[Y] \end{aligned}$$

# Properties of Covariance

- Let $X$, $Y$, and $Z$ be rv's and $a$ and $b$ scalars. Then
  (i) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
  (ii) $\text{Cov}(X, a) = 0$
  (iii) $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$
  (iv) $\text{Cov}(a + bX, Y) = b\text{Cov}(X, Y)$
  (iv) $\text{Cov}(X, X) = \text{Var}(X)$
- Covariance will also be a part of variances of sums of rvs:

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

# Correlation

- Covariance also has odd units - (units of X) x (units of Y)
- Thus, we often use the Correlation between $X$ and $Y$:

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

# Correlation

- Correlation is unitless - for any $X$ and $Y$, $|Corr(X, Y)| \in [0, 1]$
- If $Corr(X, Y) = 0$ we say the two variables are uncorrelated
- On the other hand, $Corr(X, Y) = 1$ if and only if $Y = a + bX$ for some scalar $a$ and some positive scalar $b$ (a negative $b$ implies $Corr(X, Y) = -1$)

# Conditional Distributions

- The Conditional Distribution tells us about the likelihood of a given outcome(s) for one variable if we know the outcome of another variable
- For discrete random variables $X$ and $Y$, and $x_i$ st $P(X = x_i) > 0$

$$P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_i)}{P(X = x_i)}$$

- For continuous $X$, $Y$, we'll have the conditional pdf:

$$f(y|x) = \frac{f(x,y)}{f(x)}$$

# Conditional Expectations

- Defining conditional distributions allows us to consider Conditional Expectations - our "best guess" at the value of one rv given what we know about another rv
- For discrete rv's, where $Y$ takes on $k$ values

$$E[Y|X = x_i] = \sum_{j=1}^{k} y_j P(Y = y_j|X = x_i)$$

- Continuous rv's have the definition

$$E[Y|X = x] = \int_{-\infty}^{\infty} yf(y|x)dy$$

# Conditional Expectations

- It's important to note that general conditional expectations of rv's, $E[Y|X]$ are *functions*
- A conditional expectation evaluated at a specific value of the conditioning variable, $E[Y|X = x]$ is a number - we can solve it for a specific value given the formulae on the previous slide
- However, because those formulas spit out different values for different $x$, the general form $E[Y|X]$ is a function

# Properties of Conditional Expectations

- Let $X$, $Y$, and $Z$ be rv's. For any functions $g$ and $h$,
  (i) $E[g(X) + h(X)Y|X] = g(X) + h(X)E[Y|X]$;
  (ii) $E[Y + Z|X] = E[Y|X] + E[Z|X]$;
  (iii) if $Y \leq Z$ (always), then $E[Y|X] \leq E[Z|X]$.
- We can notice that these properties are all quite similar to the properties of unconditional expectations, except that functions of the conditional variable are taking the place of constants - if we evaluate these functions of $X$ at a specific $x$, those functions become constants!

# Law of Iterated Expectations

- We can go from conditional expectations to unconditional expectations using an extremely important tool - the Law of Iterated Expectations

$$E[Y] = E[E[Y|X]]$$

# Law of Iterated Expectations

- If $E[Y|X] = E[Y]$, we say that $Y$ is Mean Independent of $X$
- If $Y$ is mean independent of $X$, then
  (i) $E[YX] = E[Y]E[X]$;
  (ii) $\text{Corr}[Y, X] = 0$.
- Because mean independence implies uncorrelatedness, but not necessarily vice versa, mean independence can be said to be "stronger" than uncorrelatedness

# Conditional Variance

- Just as we can think about the expectation of a variable conditioning on the value of another variable, we can do the same with variance. The Conditional Variance of $Y$ given $X$ is:

$$\begin{aligned} Var(Y|X) &= E[(Y - E[Y|X])^2|X] \\ &= E[Y^2 - 2YE[Y|X] + E[Y|X]^2|X] \\ &= E[Y^2|X] - 2E[Y|X]E[Y|X] + E[Y|X]^2 \\ \Rightarrow Var(Y|X) &= E[Y^2|X] - E[Y|X]^2 \end{aligned}$$

- Note again that this is a function of $X$

# Properties of Conditional Variance

- Let $X$ and $Y$ be rv's. For any functions $g$ and $h$,

$$\text{Var}[g(X) + h(X)Y|X] = h^2(X)\text{Var}[Y|X]$$

- Similar to unconditional variance with functions of $X$ taking place of scalar - again, for a specific value $x$, those functions become scalars
- There is also the Law of Total Variance, which states:

$$Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$$

# Independence

- Rv's $X$ and $Y$ are Independent, denoted $X \perp Y$, if, for any sets $A$ and $B$:

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

# Independence

- For discrete $X$ and $Y$ we observe that:

$$P(Y = y_j | X = x_j) = \frac{P(Y = y_j, X = x_j)}{P(X = x_j)}$$
$$= \frac{P(Y = y_j)P(X = x_j)}{P(X = x_j)}$$
$$= P(Y = y_j)$$

- Similarly, for continuous $X$ and $Y$, we have that

$$f(y|x) = f(y)$$

- Under independence, conditional distributions are the same as the unconditional distribution - the value of $X$ conveys no "information" about $Y$

# Independence

- The coincidence of the conditional and unconditional distributions implies mean independence
- Thus, it is the case that independence $\Rightarrow$ mean independence $\Rightarrow$ uncorrelatedness
- However, the reverse of the above is not necessarily true. Thus, we have a hierarchy of notions of variables being unrelated

# The Normal Distribution

- $X$ is Normally Distributed with mean $\mu$ and variance $\sigma^2$ if it has the pdf

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-1}{2}(\frac{x-\mu}{\sigma})^2}$$

- We denote normally distributed variables as $X \sim N(\mu, \sigma^2)$
- This pdf produces a typical "Bell-curve" like distribution, which characterizes many naturally-occurring distributions - height, standardized test scores, shoe size...

# Standard Normal Distribution

- The special case of $\mu = 0$ and $\sigma^2 = 1$ is called the Standard Normal. The cdf of a standard normal is denoted by $\Phi(x)$
- Any normal distribution can be "standardized" by taking the transformation $\frac{X-\mu}{\sigma}$. Thus, if $X \sim N(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0,1)$
- The standard normal has a nice interpretation - a value of 1 indicates 1 std dev away from the mean (2 is 2 std dev's, -1 is -1 std dev's, etc)

# Properties of Normal Distribution

- If $X_1$, $X_2$, $X_3$,... are independent normal rv's and $a_1$, $a_2$, $a_3$,... are scalars, then $\sum_{i=1}^{m} a_i X_i$ is also normal with the distribution:

$$\sum_{i=1}^{m} a_i X_i \sim N(\sum_{i=1}^{m} a_i \mu_i, \sum_{i=1}^{m} a_i^2 \sigma_i^2)$$

- The normal distribution is also symmetric around its mean. Thus, for the standard normal, it is the case that:

$$\Phi(x) = 1 - \Phi(-x)$$