THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 321: Introduction to Probability and Statistics
Review: Continuous r.v. review

Purnamrita Sarkar
Department of Statistics and Data Sciences
The University of Texas at Austin

Random variables

- If the range of our random variable X is finite or countable, we call it a **discrete random variable**.
- We can write the probability that the random variable X takes on a specific value x using the **probability mass function**, $p_X(x) = \mathbf{P}(X = x)$.

- If the range of X is uncountable, and no single outcome x has $\mathbf{P}(X = x) > 0$, we call it a **continuous random variable**.
- Because $\mathbf{P}(X = x) = 0$ for all X, we can't use a PMF.
- However, for any *range* B of X – e.g. $B = \{x | x < 0\}$, $B = \{|3 \leq x \leq 4\}$ – we have $\mathbf{P}(X \in B) \geq 0$.
- We can define the **probability density function** $f_X(x)$ as the non-negative function such that

$$\mathbf{P}(X \in B) = \int_B f_X(x)dx$$

for all subsets B of the line.

Cumulative distribution functions

- For a **discrete** random variable, to get the probability of X being in a range B, we **sum** the PMF over that range:

$$\mathbf{P}(X \in B) = \sum_{x \in B} p_X(x)$$

e.g. if $X \sim Binomial(10, 0.2)$

$$\mathbf{P}(2 < X \leq 5) = p_X(3) + p_X(4) + p_X(5) = \sum_{k=3}^{5} \binom{10}{k} 0.2^k (1 - 0.2)^{10-k}$$

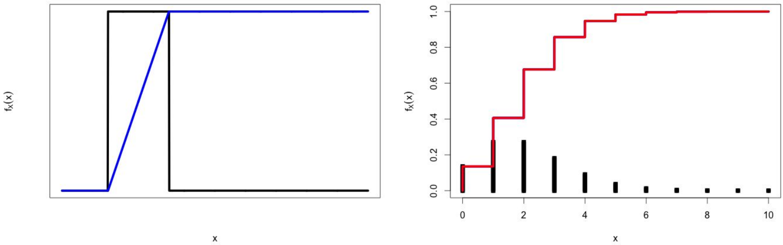
- For a **continuous** random variable, to get the probability of X being in a range B, we **integrate** the PDF over that range:

$$\mathbf{P}(X \in B) = \int_B f_X(x) dx$$

$$\mathbf{P}(2 < X \leq 5) = \int_2^5 f_X(x) dx$$

Cumulative distribution functions

- In both cases, we call the probability $\mathbf{P}(X \leq x)$ the **cumulative distribution function** (CDF) $F_X(x)$ of X



Common discrete random variables

We have looked at four main types of discrete random variable:

- **Bernoulli**: We have a biased coin with probability of heads p. A Bernoulli random variable is 1 if we get heads, 0 if we get tails.
  - If $X \sim \textit{Bernoulli}(p)$, $p_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{otherwise.} \end{cases}$
  - Examples: Has disease, hits target.
- **Binomial**: We have a sequence of n biased coin flips, each with probability of heads p – i.e. a sequence of n independent $\textit{Bernoulli}(p)$ trials. A Binomial random variable returns the number of heads.
  - If $X \sim \textit{Binomial}(n, p)$, $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$
  - p^k because we have heads (prob. p) k times, $(1-p)^{n-k}$ because we have tails (prob. $1-p$) $n-k$ times.
  - Why $\binom{n}{k}$? Because this is the number of sequences of length n that have exactly k heads.
  - Examples: How many people will vote for a candidate, how many of my seeds will sprout.

Common discrete random variables

- **Geometric**: We have a biased coin with probability of heads p. A geometric random variable returns the number of times we have to throw the coin before we get heads. e.g. If our sequence is $(T, T, H, T, \dots)$, then $X = 3$.
  - If $X \sim Geometric(p)$, $p_X(k) = (1-p)^{k-1}p$
  - Prob. of getting a sequence of $k-1$ tails and then a head.
  - $E[X] = 1/p$ (should know this) and $\text{var}(X) = (1-p)/p^2$ (don't need to know this)
  - $P(X > k) = 1 - P(X \leq k) = (1-p)^k$
  - Memoryless property: $P(X > k + j | X > j) = P(X > k)$
- **Poisson**: Independent events occur, on average, λ times over a given period/distance/area. A Poisson random variable returns the number of times they actually happen.
  - If $X \sim Poisson(\lambda)$, $p_X(k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$.
  - $E[X] = \lambda$ and $\text{var}(X) = \lambda$.
- The Poisson distribution with $\lambda = np$ is a good approximation to the $Binomial(n, p)$ distribution, when n is large and p is small.

Common continuous random variables

- **Uniform** random variable: X takes on a value between a lower bound a and an upper bound b, with all values being equally likely.
  - If $X \sim Uniform(a, b)$, $f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \leq X \leq b \\ 0 & \text{otherwise.} \end{cases}$
- $E[X] = (a+b)/2$. $\text{var}(X) = (b-a)^2/12$
- **Normal** random variable: X follows a bell-shaped curve with mean μ and variance σ^2. No upper or lower bound.
  - If $X \sim Normal(\mu, \sigma^2)$, then $f_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- If $X \sim Normal(\mu, \sigma^2)$, then $Z = \dfrac{X-\mu}{\sigma} \sim Normal(0, 1)$
- $E[X] = \mu$ and $\text{var}(X) = \sigma^2$.
- **Exponential** random variable: X takes non-negative values.
  - If $X \sim Exponential(\lambda)$, $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$
- $E[X] = 1/\lambda$, $\text{var}(X) = 1/\lambda^2$.
- Always remember, whenever you have an integration of the form $\int_0^\infty \lambda x \exp(-\lambda x) dx$, you should be able to go around partial integration by using the expectation formula of an Exponential.

The standard normal

- The pdf of a standard normal is defined as:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

- The CDF of the standard normal is denoted Φ:

$$\Phi(z) = \mathbf{P}(Z \leq z) = \mathbf{P}(Z < z) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{z} e^{-t^2/2} dt$$

- We cannot calculate this analytically.
- The **standard normal table** lets us look up values of $\Phi(y)$ for $y \geq 0$

	.00	.01	.02	0.03	0.04	⋯
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	⋯
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	⋯
0.2	0.5793	**0.5832**	0.5871	0.5910	0.5948	⋯
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	⋯
⋮	⋮	⋮	⋮	⋮	⋮	

$$\mathbf{P}(Z < 0.21) = 0.5832$$

CDF of a normal random variable

The pdf of a $X \sim N(\mu, \sigma^2)$ r.v. is defined as:

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

If $X \sim N(3, 4)$, what is $\mathbf{P}(X < 0)$?

- First we need to **standardize**:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 3}{2}$$

- So, a value of $x = 0$ corresponds to a value of $z = -1.5$
- Now, we can translate our question into the standard normal:

$$\mathbf{P}(X < 0) = \mathbf{P}(Z < -1.5) = \mathbf{P}(Z \leq -1.5)$$

- Problem... our table only gives $\Phi(z) = \mathbf{P}(Z \leq z)$ for $z \geq 0$.
- But, $\mathbf{P}(Z \leq -1.5) = \mathbf{P}(Z \geq 1.5)$, due to symmetry.
- Our table only gives us "less than" values.
- But, $\mathbf{P}(Z \geq 1.5) = 1 - \mathbf{P}(Z < 1.5) = 1 - \mathbf{P}(Z \leq 1.5) = 1 - \Phi(1.5)$.
- And we're done!
  $\mathbf{P}(X < 0) = 1 - \Phi(1.5) = $ (look at the table...) $1 - 0.9332 = 0.0668$

Multiple random variables

- We can have *multiple* random variables associated with the same sample space.
- e.g. if our experiment is an infinite sequence of coin tosses, we might have:
  - X = number of heads in the first 10 coin tosses.
  - Y = outcome of 3rd coin toss.
  - Z = number of coin tosses until we get a head.
- e.g. if our experiment is to pick a random individual, we might have:
  - X = age.
  - $Y = 1$ if male, 0 otherwise.
  - Z = height.
- We can look at the probability of outcomes of a single random variable in isolation:
  - Discrete case: The **marginal PMF** of X is just $p_X(x) = \mathbf{P}(X = x)$ – the PMF associated with X in isolation.
  - Continuous case: The **marginal PDF** of X is the function $f_X(x)$ such that $\mathbf{P}(X \in B) = \int_B f_X(x)dx$.

Multiple random variables

- We can look at the probability of two (or more) random variables' outcomes together.
- For example, if we are interested in $\mathbf{P}(X < 3)$ and $\mathbf{P}(2 < Y \leq 5)$, we might want to know $\mathbf{P}((X < 3) \cap (2 < Y \leq 5))$ – the probability that both events occur.
  - Discrete case: The **joint PMF**, $p_{X,Y}(x, y)$, of X and Y gives probability that X and Y both take on specific values x and y, i.e.

$$p_{X,Y}(x, y) = \mathbf{P}(\{X = x\} \cap \{Y = y\})$$

  - Continuous case: The **joint PDF**, $f_{X,Y}(x, y)$, of X and Y is the function we integrate over to get $\mathbf{P}((X, Y) \in B)$, i.e.

$$\mathbf{P}((X, Y) \in B) = \iint_{x,y \in B} f_{X,Y}(x, y) dx\, dy$$

Multiple random variables

- We can get from the **joint PMF** of X and Y to the **marginal PMF** of X by summing over (marginalizing over) Y:

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

- We can get from the **joint PDF** of X and Y to the **marginal PDF** of X by integrating over (marginalizing over) Y:

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

Multiple random variables

- If we have two events A and B, if we *know* B is true, we can look at the **conditional probability** of A given B, $\mathbf{P}(A|B)$
  - We know that $\mathbf{P}(A|B) = \frac{\mathbf{P}(A,B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B|A)\mathbf{P}(A)}{\mathbf{P}(B)}$
- If we have a **discrete** random variable X and an event B, if we *know* B is true, the **conditional PMF** of X given B gives the PMF of X given B:

$$p_{X|B}(x) = \mathbf{P}(\{X = x\}|B)$$

- If we have a **discrete** random variable X and a **discrete** random variable Y, the **conditional PMF** of X given Y gives the PMF of X given $Y = y$:

$$p_{X|Y}(x|y) = \mathbf{P}(\{X = x\}|\{Y = y\})$$

- Since $\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B|A)\mathbf{P}(A)}{\mathbf{P}(B)}$, we know that

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

Multiple random variables

- If we have a **continuous** random variable X and an event B, if we *know* B is true, the **conditional PDF** of X given B is the function $f_{X|B}(x)$ such that

$$\mathbf{P}(X \in A|B) = \int_A f_{X|B}(x)dx$$

- If the event B corresponds to a range of outcomes of X, we have

$$\begin{aligned}\mathbf{P}(X \in A|X \in B) &= \frac{\mathbf{P}(\{X \in A\} \cap \{X \in B\})}{\mathbf{P}(X \in B)} = \frac{\int_{A \cap B} f_X(x)dx}{\mathbf{P}(X \in B)} \\ &= \int_A f_{X|\{X \in B\}}(x)dx\end{aligned}$$

- Therefore, $f_{X|\{X \in B\}} = \begin{cases} \dfrac{f_X(x)}{\mathbf{P}(X \in B)} & \text{if } x \in B \\ 0 & \text{otherwise} \end{cases}$

Multiple random variables

- If the event B corresponds to the outcome y of another random variable Y, the **conditional PDF** of X given $Y = y$ is the function $f_{X|Y}(x|y)$ such that

$$\mathbf{P}(\{X \in A\}|Y = y) = \int_A f_{X|Y}(x|y)dx$$

- If Y is a continuous random variable, we have the relationship

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

- If Y is a discrete random variable, we have the relationship

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

Expectation and Variance

- The **expectation** (expected value, mean) $E[X]$ of a random variable X is the number we expect to get out, on average, if we repeat the experiment infinitely many times.
- The **variance** of a random variable is the expected value of $(X - E[X])^2$. It is a measure of how far away from our expectation we expect to be.
- If we know the PMF or PDF of a random variable, we can calculate its mean and variance.

Discrete:

$$E[X] = \sum_x x p_X(x)$$
$$\text{var}(X) = \sum_x (x - E[X])^2 p_X(x)$$
$$= E[X^2] - E[X]^2$$

Continuous:

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$
$$\text{var}(X) = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx$$
$$= E[X^2] - E[X]^2$$

Expectation and Variance

- A random variable is a mapping from our sample space to the real line.
- So, any function of a random variable is itself a random variable.
- So, we can calculate its mean and expectation!

Discrete:

$$E[g(X)] = \sum_x g(x) p_X(x)$$

$$\text{e.g. } E[X^2] = \sum_x x^2 p_X(x)$$

$$\begin{aligned} \text{var}(g(X)) &= \sum_x (g(x) - E[g(X)])^2 p_X(x) \\ &= E[g(X)^2] - E[g(X)]^2 \end{aligned}$$

Continuous:

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

$$\text{e.g. } E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx$$

$$\begin{aligned} \text{var}(g(X)) &= \int_{-\infty}^{\infty} (g(x) - E[g(X)])^2 f_X(x) dx \\ &= E[g(X)^2] - E[g(X)]^2 \end{aligned}$$

- If $g(X) = aX + b$, then we have:

Discrete:

$$\begin{aligned} E[aX + b] &= aE[X] + b \\ \text{var}(aX + b) &= a^2 \text{var}(X) \end{aligned}$$

Continuous:

$$\begin{aligned} E[aX + b] &= aE[X] + b \\ \text{var}(aX + b) &= a^2 \text{var}(X) \end{aligned}$$

Expectation and Variance

- If X and Y are two random variables, any function of X and Y is also a random variable.
- So, we can look at the expectation and variance of $g(X, Y)$

Discrete:

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) p_{X,Y}(x,y)$$

$$\begin{aligned}\text{var}(g(X,Y)) &= \sum_x \sum_y (g(x,y) - E[g(X,Y)])^2 p_{X,Y}(x,y) \\ &= E[g(X,Y)^2] - E[g(X,Y)]^2\end{aligned}$$

Continuous:

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

$$\begin{aligned}\text{var}(g(X,Y)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (g(x,y) - E[g(X,Y)])^2 f_{X,Y}(x,y) dx\, dy \\ &= E[g(X,Y)^2] - E[g(X,Y)]^2\end{aligned}$$

Linearity of expectation

- If $g(X, Y) = aX + bY + c$, we observe **linearity of expectation**:

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

$$E[aX^2 + bY^2 + c] = aE[X^2] + bE[Y^2] + c$$

- $\text{var}(aX + bY + c) = a^2\text{var}(X) + b^2\text{var}(Y)$ **Only if X and Y are independent!**
- When X and Y are dependent, then there are covariance terms. See probability review after the midterm.
- Adding a constant never changes the variance.
- If you have the mean and the variance, you can easily calculate $E[X^2]$ by using the formula

$$\text{var}(X) = E[X^2] - E[X]^2$$

Conditional expectation

- We can also look at the **conditional** expectation of X, given some event Y.
- This is just the expectation of the appropriate conditional distribution!

Discrete:

$$E[X|A] = \sum_x x p_{X|A}(x)$$

Continuous:

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

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

- If $B_1, B_2, \ldots, B_n$ is a partition of Ω, we can use the **total expectation theorem** to get $E[X]$ from the conditional expectations $E[X|B_i]$:

$$E[X] = \sum_{i=1}^{n} E[X|B_i]\mathbf{P}(B_i)$$

- For a continuous random variable, if we know the conditional expectations $E[X|Y=y]$, we have

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

Independent events

- We say two events A and B are **independent** if $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.
- If $\mathbf{P}(B) > 0$, this implies $\mathbf{P}(A|B) = \mathbf{P}(A)$ – i.e. knowing B tells us nothing about A.
- We say more than two events $A_1, A_2, \ldots, A_n$ are independent if, for any subset S of $\{1, 2, \ldots, n\}$, $\mathbf{P}\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \mathbf{P}(A_i)$
- We say more than two events $A_1, A_2, \ldots, A_n$ are *pairwise* independent if, for any pair A_i, A_j, $\mathbf{P}(A_i \cap A_j) = \mathbf{P}(A_i)\mathbf{P}(A_j)$

Conditionally independent events

- We say any two events A and B are **conditionally independent** given some third event C if

$$\mathbf{P}(A \cap B|C) = \mathbf{P}(A|C)\mathbf{P}(B|C)$$

- If $\mathbf{P}(B|C) > 0$, this implies $\mathbf{P}(A|B, C) = \mathbf{P}(A|C)$ – i.e. if we already know C, knowing B tells us nothing more about A.
- *Conditional independence does not imply independence!*
- *Independence does not imply conditional independence!*

Independent random variables

- We can extend this concept to random variables.
- We say two random variables X and Y are independent if, for all x and y,

Discrete:

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

Continuous:

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

- If $p_Y(y)$ (discrete case)/$f_Y(y)$ (continuous case is zero, this implies

Discrete:

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

Continuous:

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

– i.e. knowing y tells us nothing about X (and vice versa)!

Expectation and Variance of independent random variables

- If X and Y are independent random variables, $p_{X,Y}(x,y) = p_X(x)p_Y(y) / f_{X,Y}(x,y) = f_X(x)f_Y(y)$, so we have

Discrete:

$$
\begin{aligned}
E[XY] &= \sum_{x,y} xy p_{X,Y}(x,y) \\
&= \sum_{x} x p_X(x) \sum_{Y} y p_Y(y) \\
&= E[X]E[Y]
\end{aligned}
$$

Continuous:

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

- We also have:

$$
\begin{aligned}
\text{var}(X+Y) &= E[(X+Y)^2] - E[X+Y]^2 \\
&= E[X^2] + 2E[XY] + E[Y^2] - 2E[X]^2 - 2E[X]E[Y] - 2E[Y^2] \\
&= \text{var}(X) + \text{var}(Y)
\end{aligned}
$$

Conditional PDF and conditional expectation

- We all know how to calculate expectation of a continuous or discrete random variable
  - e.g. $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$
- ... and we're generally OK with calculating the expectation of a function of a random variable
  - e.g. $E[X^3] = \int_{-\infty}^{\infty} x^3 f_X(x) dx$
- If we condition on some separate event, it's still pretty straightforward.
- The conditional PDF of X given some event A describes the probability distribution of X given some event A.
  - e.g. X might be "height" and A might be "is male"; $f_{X|A}(x)$ is the PDF of a man's height.
  - X might be "weight" and A might be "$X > 140lb$"; $f_{X|X>140}(x)$ is the PDF of weight, given weight$>$140.
- The event might be the outcome of another random variable.
  - e.g. e.g. X might be height, and Y might be age; $p_{X|Y}(x|y)$ is the PDF of someone's height given their age is y.
- In all these cases, we can calculate the *conditional expectation* and *conditional variance* of X.

Law of total expectation

- Let's look at a concrete example. Let's say the distribution of heights for men is a normal distribution with mean 69 and variance 9, and the distribution of heights for women is a normal distribution with mean 64 and variance 9.
- What is the conditional expectation of height, given a subject is male?
- $E[H|M] = 69$" - it's the mean of a $Normal(69, 9)$ distribution.
- Similarly, the conditional expectation of height given a subject is female is 64".
- If a population contains equal numbers of men and women, what is the overall expected height?
- $E[H] = 66.5''$... why?

Law of total expectation

- Well, we have a 0.5 probability of being male, and if we're male the conditional expectation is 69... we have a 0.5 probability of being female, and then the conditional expectation is 64.

$$\underbrace{E[H] = E[H|M]\mathbf{P}(M) + E[H|F]\mathbf{P}(F)}_{\textbf{law of total expectation}} = 69 \times 0.5 + 64 \times 0.5 = 66.5$$

- What if our population were 2/3 men? What would the total expectation be then?

$$E[H] = E[H|M]\mathbf{P}(M) + E[H|F]\mathbf{P}(F) = 69 \times \frac{2}{3} + 64 \times \frac{1}{3} = 67.33$$

Conditional expectation: Conditioning on a concrete event

- Let's look at another example
- A bird leaves its nest heading north, at 1pm. Let X be the speed at which a bird flies, and Y be the bird's distance north from the nest at 1:30pm.
- Assume $X \sim \mathit{Uniform}(3, 6)$ and $Y|X = x \sim \mathit{Normal}(x/2, 1)$
- Let's say we now discover the bird's flight speed is 4mph. What is the conditional expectation of its distance?
- If its speed is 4mph, the distance is a normal random variable with mean 2 miles and variance 1. So, $E[Y|X = 4] = 2$.
- Note that this **doesn't depend on the distribution of X!** We are conditioning on a fixed value of $X = x$, so we don't need to incorporate its distribution.
- What is the conditional variance of Y, given $X = 4$?
- $\text{var}(Y|X = 4) = 1$.

Conditional expectation: Conditioning on a random event

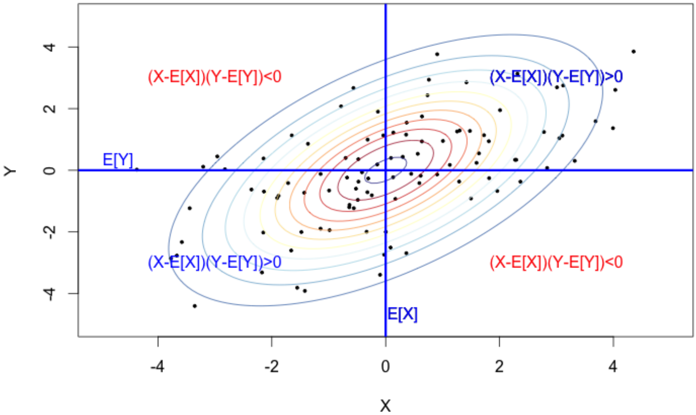
- What if we don't know the bird's speed? What can we say about the conditional expectation of the distance, given this unknown speed?
- Well, the conditional expectation will always be half the speed!
- If we know $X = x$, then $E[Y|X = x] = x/2$. It is a **function of x!**
- Since X is random, the conditional expectation of Y given X, $E[Y|X]$ is random.
- Note we are now conditioning on a **random variable** rather than a **specific set**.
- We know that $E[Y|X] = X/2$... i.e. the conditional expectation of Y given X is a $Uniform(1.5, 3)$ distribution.

Correlation and Covariance

- The **expectation** of a random variable tells us **where we expect it to be, on average**.
- The **variance** of a random variable tells us **how far away from the expectation** it is likely to be.
- If we have *two* random variables, we might also be interested in how much they **jointly** vary from the mean.
- We might be interested in knowing whether a larger-than-expected X comes with a larger-than-expected Y.
- The covariance between X and Y captures this:

$$\begin{aligned}\text{cov}(X, Y) =& E\left[(X - E[X])(Y - E[Y])\right] \\ =& E[XY] - E[X]E[Y]\end{aligned}$$

Covariance



- A **positive** covariance means that, if $X > E[X]$, we are likely to have $Y > E[Y]$
- A **negative** covariance means that, if $X > E[X]$, we are likely to have $Y < E[Y]$.

Covariance: Example

A die is rolled twice. Let X be the sum of outcomes and Y be the first outcome minus the second. Calculate $\text{cov}(X, Y)$

- Let X_1 be the first outcome and X_2 the second.
- $X = X_1 + X_2$ and $Y = X_1 - X_2$
- $\text{cov}(X_1 + X_2, X_1 - X_2) =$ $\text{cov}(X_1, X_1) - \text{cov}(X_1, X_2) + \text{cov}(X_1, X_2) - \text{cov}(X_2, X_2) = 0$, since $\text{cov}(X_1, X_1) = \text{cov}(X_2, X_2) = \text{var}(X_1)$.

Covariance: A slightly harder example

Let X be the number of ones and Y be the number of 2's that occur in n rolls of a fair die. Calculate $\text{cov}(X, Y)$

- Let $X_i = 1$ if the i^{th} throw yields a 1 and $Y_i = 1$ if the i^{th} throw yields a 2.
- We want $\text{cov}(\sum_i X_i, \sum_j Y_j)$
- This equals $\sum_{ij} \text{cov}(X_i, Y_j)$
- But X_i and Y_j are independent when $i \neq j$, and so $\text{cov}(X_i, Y_j) = 0$ when $i \neq j$
- So $\text{cov}(X, Y) = \sum_i \text{cov}(X_i, Y_i) = \sum_i (E[X_i Y_i] - E[X_i]E[Y_j])$.
  However, $E[X_i Y_i] = P(X_i = 1, Y_i = 1) = 0$ and so the answer is $n(0 - 1/36) = -n/36$

Correlation

- The covariance can be hard to interpret.
- The sign gives us the direction of the relationship.
- The size depends on **both** the strength of the relationship **and** how far X and Y tend to be from their mean – i.e. the variances of X and Y.
- The **correlation coefficient** $\rho_{X,Y}$ is a standardized version of the covariance.

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

- It takes away the effect of the variances, so we always have $-1 \leq \rho_{X,Y} \leq 1$
  - $\rho = 0$ implies zero covariance.
  - $|\rho| = 1$ iff there is a linear relationship between X and Y.

Covariance and correlation properties

- $\text{cov}(X, X) = \text{var}(X)$
- $\text{cov}(X, Y) = \text{cov}(Y, X)$
- $\text{cov}(X, a) = 0$ for some constant a.
- $\text{cov}(aX, bY) = ab\text{cov}(X, Y)$
- $\text{cov}(X, b + Y) = \text{cov}(X, Y)$
- If $X = aY + b$ and $a > 0$ then $\rho_{X,Y} = 1$
- If $X = aY + b$ and $a < 0$ then $\rho_{X,Y} = -1$

Correlation: Example

- We just calculated the **correlation** between the number of ones and the number of twos that occur in n rolls of a fair die: $\mathrm{cov}(X, Y) = -n/36$.
- What is the correlation?
- Well, we first need to know the variances.
- X is a binomial random variable... each roll comes up ones with probability $1/6$.
- So, we know $\mathrm{var}(X) = 5n/36$.
- Similarly, $\mathrm{var}(Y) = 5n/36$.
- So,

$$\rho_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}} = \frac{-n/36}{5n/36} = -\frac{1}{5}$$

What you should know

- Independence and conditional independence. The difference.
- What is a valid (joint) pdf ?
- How to calculate mean, variance, marginal pdf of a random variable.
- How to look up a normal table?
- Forms of the common distributions and at least their means. Probably a good idea to keep the variances on a cheat sheet.
- Bayes rule.
- Conditional probabilities and expectations. Law of total probability and total expectation theorem
- How to calculate covariance and correlation. Rules of covariance and correlation.