STA257: Probability and Statistics 1

Instructor: Katherine Daignault

Department of Statistical Sciences
University of Toronto

Week 9

Outline

Outline

Calculus Refresher: Definite Integrals

- Again you will need to be very comfortable with taking definite integrals of functions.
- Definition: $\int_a^b f(x)dx = F(x)|_a^b = F(b) - F(a)$
- Some useful properties:
  - Reversing limits: $\int_a^b f(x)dx = -\int_b^a f(x)dx$
  - Additivity: $\int_a^b [f(x) + g(x)]dx = \int_a^b f(x)dx + \int_a^b g(x)dx$
- Some useful results (written as indefinite integrals):
  - $\int adx = ax + C$
  - $\int x^a dx = \frac{x^{a+1}}{a+1} + C$
  - $\int (1/x)dx = ln(x) + C$
  - $\int e^{ax} dx = (1/a)e^{ax} + C$
  - $\int a^x dx = a^x / ln(a) + C$
  - $\int ln(x)dx = xln(x) - x + C$

Calculus Review - Multiple Integration

- Just as we need to take partial derivatives for multiple variables, we need to take multiple integrals over all the variables in our function.
- We define the double integral over $f(x, y)$ as

$$\int_{supp(X)} \int_{supp(Y)} f(x, y) dy dx$$

  where we use the term $supp(X)$ to denote the support of X, or the values that X is defined over.
- In this case, you (1) take the integral with respect to Y, treating the x's as constants, then (2) take the integral of the result with respect to X.

Calculus Review - Multiple Integration

- Be careful about the order of the integrals!
  - If you want to integrate Y first, then the inner integral must correspond to the integral over y values
  - If you want to integrate X first, then you need to switch the order of the integrals to

$$\int_{supp(Y)} \int_{supp(X)} f(x,y)dxdy$$

- Since a single integral is the area under a curve, a double integral find the area **between** two curves.
- Often the support of one variable will be defined in terms of the other variable so you will need to be able to manipulate the bounds of integration.

Calculus Review - Multiple Integration

- If we are dealing with a function $f(x,y)$ defined on the region

$$F = \{(x,y) : a \leq x \leq b, p(x) \leq y \leq q(x)\}$$

then we must take the integral in the following order

$$\int\int_F f(x,y)dA = \int_a^b \int_{p(x)}^{q(x)} f(x,y)dydx$$

- If I wanted to take the integral in the reverse order, then I would need to determine the functions $r(y)$ and $s(y)$ so that

$$\int\int_F f(x,y)dA = \int_c^d \int_{r(y)}^{s(y)} f(x,y)dxdy$$

and $F = \{(x,y) : c \leq y \leq d, r(y) \leq x \leq s(y)\}$

Outline

Conditional Probability

- Recall from Week 2 that when two (or multiple) events are not independent, then we may consider the conditional probability of one given another.
- We denoted the conditional probability of A given B as
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$
- Since random variables work in much the same way as events, we can define the conditional probability of random variable X given random variable Y

Discrete Conditional Distributions

- As we are no longer talking about events, we must adapt the definition of conditional probability for use with distribution/density functions.
- Consider two random variables X and Y that are jointly distributed discrete variables, with joint PMF

$$p_{XY}(x, y) = P(X = x, Y = Y)$$

  and marginal PMFs $p_X(x) = P(X = x)$ and $p_Y(y) = P(Y = y)$.
- Then the conditional PMF of X given Y is

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

  as long as $P(Y = y_j) > 0$

Discrete Conditional Distributions

- We can use the shorthand $p_{X|Y}(x \mid y) = \frac{p_{XY}(x,y)}{p_Y(y)}$ to refer to this conditional distribution.
- It can be shown that $p_{X|Y}(x \mid y$ is a valid probability distribution function since it satisfies the criteria:
  - it is non-negative (i.e. $p_{X|Y}(x \mid y) \geq 0$ for all values of x, y)
  - it sums to 1
- We also have the analogous result as with events to determine if X and Y are independent:

$$p_{X|Y}(x \mid y) = p_X(x) \text{ and } p_{Y|X}(y \mid x) = p_Y(y)$$

if X and Y are independent

Example: Coin tossing example from Week 8

- Suppose we toss a fair coin 3 times.
  - X denotes the number of heads on the first toss
  - Y denotes the total number of heads.
- We can represent the joint probabilities that $(X = x_i, Y = y_j)$ using a contingency table:

	y			
x	0	1	2	3
0	$1/8$	$2/8$	$1/8$	0
1	0	$1/8$	$2/8$	$1/8$

- We had found that the marginal PMFs for this example were

$$p_Y(y) = \begin{cases} 1/8, & y = 0,3 \\ 3/8 & y = 1,2 \end{cases} \text{ and } p_X(x) = \begin{cases} 1/2, & x = 0 \\ 1/2, & x = 1 \end{cases}$$

Example: Coin tossing example from Week 8 (cont.)

- We can use these to find the conditional PMFs given that $Y = 1$.
- From the definition,

$$P(X = x_i \mid Y = 1) = \frac{P(X = x_i, Y = 1)}{P(Y = 1)}$$

  so to get the distribution, I need to calculate this for each value of X.
- When $X = 0$:

$$P(X = 0 \mid Y = 1) = \frac{2/8}{3/8} = \frac{2}{3}$$

- When $X = 1$:

$$P(X = 1 \mid Y = 1) = \frac{1/8}{3/8} = \frac{1}{3}$$

Example: Coin tossing example from Week 8 (cont.)

- We can of course find conditional PMFs given e.g. $X = 1$.
- From the definition,

$$P(Y = y \mid X = 1) = \frac{P(X = 1, Y = y_j)}{P(X = 1)}$$

- Now we have to find 4 conditional probabilities:
  - When $Y = 0$: $P(Y = 0 \mid X = 1) = \frac{0}{1/2} = 0$
  - When $Y = 1$: $P(Y = 1 \mid X = 1) = \frac{1/8}{1/2} = \frac{1}{4}$
  - When $Y = 2$: $P(Y = 2 \mid X = 1) = \frac{2/8}{1/2} = \frac{1}{2}$
  - When $Y = 1$: $P(Y = 3 \mid X = 1) = \frac{1/8}{1/2} = \frac{1}{4}$

Example: Coin tossing example from Week 8 (cont.)

- Suppose now we are asked to calculate a specific probability.
- Let's find the probability that we will get 2 or more total heads given that we know we got a head on the first toss.
- This corresponds to finding the probability $P(Y \geq 2 \mid X = 1)$
- But we just found $p(Y \mid X = 1)$, so this means we just sum up over the relevant probabilities from this conditional PMF:

$$P(Y \geq 2 \mid X = 1) = P(Y = 2 \mid X = 1) + P(Y = 3 \mid X = 1)$$

- We found these on the previous slide, so we get

$$P(Y \geq 2 \mid X = 1) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$$

Exercise - Give it a try!

Recall we had 3 checkout counters, two customers that pick a counter at random. Let X denote the number of customers who choose counter 1, and Y the number of customers who choose counter 2. The joint and marginal PMFs are below. What is the conditional PMF of Y given $X = 1$?

	y			
x	0	1	2	$p_X(x)$
0	$1/9$	$2/9$	$1/9$	$4/9$
1	$2/9$	$2/9$	0	$4/9$
2	$1/9$	0	0	$1/9$
$p_Y(y)$	$4/9$	$4/9$	$1/9$	

Conditional Distributions and Multiplication Law

- As with events, we may use the conditional PMF as a means of determining the joint PMF.
- This is because we have a corresponding multiplication law:

$$p_{XY}(x, y) = p_{X|Y}(x \mid y) p_Y(y)$$

  which we get by rearranging the definition of the conditional PMF.
- We now have a very useful relationship between conditional and joint PMFs.
- Further, this relationship provides a useful application of the law of total probability:

$$p_x(x) = \sum_y p_{XY}(x, y) = \sum_y p_{X|Y}(x \mid y) p_Y(y)$$

Example: Particle Counter

Suppose a particle counter is imperfect and independently detects an incoming particle with probability p. If the distribution of the number of incoming particles in a unit of time is a $Poi(\lambda)$, what is the distribution of the number of counted particles?

- It is important to realize that we are dealing with two random variables here:
  - N, the actual/true number of particles, $N \sim Poi(\lambda)$
  - X, the number of counted particles, which sounds a little like a Binomial
- The trick is to realize that the number of counted particles X changes depending on how many particles there actually are, N
  - so we are actually given that $X \mid N = n \sim Bin(n, p)$
- What the question wants us to find is the marginal PMF for X, $p_X(x)$

Example: Particle Counter (cont.)

- So we can use our previous result about the law of total probability:

$$P(X = k) = \sum_{n=0}^{\infty} P(N = n) P(X = k \mid N = n)$$

- Next, plug in the PMFs for both X and N:

$$P(X = k) = \sum_{n=k}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \binom{n}{k} p^k (1-p)^{n-k}$$

- By turning the binomial coefficient into its factorial form and some rearranging, we have

$$P(X = k) = \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{n=k}^{\infty} \lambda^{n-k} \frac{(1-p)^{n-k}}{(n-k)!}$$

Example: Particle Counter (cont.)

- Now we use the same old trick to adjust the summation bounds:

$$P(X = k) = \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j (1-p)^j}{j!}$$

- Followed by another old trick, the series representation for e:

$$P(X = k) = \frac{(\lambda p)^k}{k!} e^{-\lambda} e^{\lambda(1-p)}$$

- Then simplify to get

$$P(X = k) = \frac{(\lambda p)^k}{k!} e^{-\lambda p}$$

which gives us that $X \sim Poi(\lambda p)$

Outline

Continuous Conditional Distributions

- The notion of conditional distributions for continuous random variables is the same as in the discrete case.
- Consider two random variables X and Y that are jointly distributed continuous variables, with joint PDF $f_{XY}(x, y)$, and marginal PDFs $f_X(x)$ and $f_Y(y)$
- Then the conditional PDF of X given Y is

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

  as long as $0 < f_Y(y) < \infty$, and 0 otherwise.
- Due to the fact that we are conditioning on $Y = y$, this means that the conditional PDF $f_{X|Y}(x \mid y)$ is a function of x only, as y has been fixed.

Example: Radioactive Particle

Suppose a radioactive particle is randomly located in a square with sides of length 1. Let X and Y denote the coordinates of the particle's location in this square. A reasonable model for the location of the particle is

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

- Suppose I want to find the conditional distribution of X given Y.
- We already found the marginals last week: $f_X(x) = f_Y(y) = 1$ when $0 \leq x \leq 1$ or $0 \leq y \leq 1$
- Therefore we can use the definition to find $f_{X|Y}(x \mid y)$

Example: Radioactive Particle (cont.)

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{1} = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

- Let's sketch out what the joint PDF actually looks like, to see what the conditional PDF of X given Y corresponds to.

Continuous Conditional Distributions

- The denominator serves to normalize the new density to have unit area (i.e. makes sure that $\int_{-\infty}^{\infty} f_{X|Y}(x \mid y)dx = 1$)
- As in the discrete case, we can write the joint PDF using a conditional PDF as

$$f_{XY}(x, y) = f_{X|Y}(x \mid y)f_Y(y)$$

  by using the multiplication law
- Further, we can find the marginal PDF of X when we have the conditional $f_{X|Y}(x \mid y)$ by

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

- We can now compute conditional probabilities.

Example: Gasoline exercise

A gas station receives a delivery of gas every week. Let X be the proportion of gasoline in the holding tank after delivery. Let Y be the proportion of gasoline in the holding tank that has been sold that week. We thus have that $Y \leq X$. If the joint PDF is $f(x, y) = 3x$, if $0 \leq y \leq x \leq 1$, find the probability that more than half the tank is sold given that $3/4$ of the tank is stocked.

- To do this, I need to find the conditional PDF of $Y \mid X$, which means I need the marginal PDF for X.

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)dy = \int_0^x 3xdy = 3x^2, \; 0 \leq x \leq 1$$

- Now I need the conditional distribution $Y \mid X$:

$$f_{Y|X}(y \mid x) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{3x}{3x^2} = \frac{1}{x}$$

Example: Gasoline exercise (cont.)

- We need to make sure that we know what values of y our conditional PDF is defined on.
- The joint distribution was defined on

$$0 \leq y \leq x \leq 1$$

- Since $f_{Y|X}(y \mid x)$ is a function of Y, for some fixed value of x, this means that $f_{Y|X}(y \mid x)$ must be defined on $0 \leq y \leq x$
- The question tells us that "given that 3/4 of the tank is stocked", so we want to find the probability being asked using the conditional distribution

$$f_{Y|X}(y \mid x = 0.75) = \frac{1}{0.75}, \; 0 \leq y \leq 0.75$$

Example: Gasoline exercise (cont.)

- The question is asking for the probability that more than half the tank is sold, given there was 75% of a tank.
- So we use the conditional PDF we just found and compute the appropriate integral:

$$P(Y \geq 1/2 \mid X = 3/4) = \int_{0.5}^{0.75} \frac{1}{0.75} dy = \frac{0.75 - 0.5}{0.75} = \frac{1}{3}$$

where we have used the new bounds for Y based on the conditional information provided.

Exercise - Give it a try!

Last week, we dealt with jointly distributed variables X and Y with joint PDF

$$f(x, y) = \begin{cases} \lambda^2 e^{-\lambda y}, & 0 \leq x \leq y \\ 0, & \text{otherwise} \end{cases}$$

where we found that $f_X(x) = \lambda e^{-\lambda x}, x \geq 0$ and $f_Y(y) = y\lambda^2 e^{-\lambda y}, y \geq 0$. Find the conditional PDF of Y given X.

Outline

Expectations of Joint Random Variables

- Expectations in the multivariate setting are basically the same as in the univariate case.
- The expectation of jointly distributed random variables again just represents the centre of the joint PDF for these variables.
- This corresponds to the values of the random variables that yield the highest probability of being observed.
- The only difference from the univariate case is that we need to sum or integrate over more than one random variable, and use a joint density/mass function

Expectations of Joint Random Variables

Expectations of Functions of Joint Random Variables

Suppose that $X_1, \dots, X_n$ are jointly distributed random variables and $Y = g(X_1, \dots, X_n)$.

1. If the X_i are discrete with mass function $p(x_1, \dots, x_n)$, then

$$E(Y) = \sum_{x_1, \dots, x_n} g(x_1, \dots, x_n) p(x_1, \dots, x_n)$$

   provided that $\sum_{x_1, \dots, x_n} \left| g(x_1, \dots, x_n) \right| p(x_1, \dots, x_n) < \infty$

2. If the X_i are continuous with joint density function $f(x_1, \dots, x_n)$, then

$$E(Y) = \int \int \cdots \int g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \cdots dx_n$$

   provided that the integral of $\left| g(x_1, \dots, x_n) \right| f(x_1, \dots, x_n) < \infty$.

Example

Let X and Y have a joint density given by

$$f_{X,Y}(x,y) = \begin{cases} 2x. & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Find $E(XY)$.

- This is just a straightforward application of the definition.
- My function g is just the product of X and Y.
- Therefore I need to integrate over both X and Y values of the joint PDF:

$$E(XY) = \int_0^1 \int_0^1 yx(2x)dxdy = \int_0^1 \frac{2y}{3}dy = \frac{1}{3}$$

Expectations of Linear Combinations of Variables

Expectation of Linear Combinations

If $X_1, \ldots, X_n$ are jointly distributed random variables with expectations $E(X_1), \ldots, E(X_n)$ and Y is a linear function of the X_i, $Y = a + \sum_{i=1}^{n} b_i x_i$, then

$$E(Y) = a + \sum_{i=1}^{n} b_i E(X_i)$$

- This follows pretty naturally from the single variable case, where $E(Y) = E(a + bX) = a + bE(X)$

Expectations of Linear Combinations of Variables

Proof

$\square$

Example: Gasoline Exercise

A gas station receives a delivery of gas every week. Let X be the proportion of gasoline in the holding tank after delivery. Let Y be the proportion of gasoline in the holding tank that has been sold that week. We thus have that $Y \leq X$. If the joint PDF is $f(x, y) = 3x$, if $0 \leq y \leq x \leq 1$, what is the expected proportion of gasoline remaining at the end of the week?

- We need to express "proportion remaining" in terms of X and Y.
- Since X is the proportion of gasoline in the tank at the beginning of the week and Y is the amount that is sold that week, we have that "proportion remaining" is $X - Y$.
- This is a linear combination of the form $aX + bY$ where $a = 1$ and $b = -1$.

Example: Gasoline Exercise (cont.)

- Using the most recent result, we see that we can write

$$E(X - Y) = E(X) - E(Y)$$

so we only need to find $E(X)$ and $E(Y)$ from the joint PDF.

- We start with

$$E(X) = \int_0^1 \int_0^x x(3x)dydx = \int_0^1 3x^3 dx = \frac{3}{4}$$

- Next find

$$E(Y) = \int_0^1 \int_0^x y(3x)dydx = \int_0^1 \frac{3x^3}{2} dx = \frac{3}{8}$$

- Finally we use the result from before to find

$$E(X - Y) = E(X) - E(Y) = \frac{3}{4} - \frac{3}{8} = \frac{3}{8}$$

Exercise - Give it a try!

A process for producing an industrial chemical yields a product containing two types of impurities. For a specific sample, let X denote the proportion of impurities in the sample, and Y denote the proportion of type 1 impurities among all impurities found. The joint distribution of X and Y is

$$f_{X,Y}(x, y) = \begin{cases} 2(1-x), & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the expected value of the proportion of type 1 impurities in the sample.

Conditional Expectations

- We have discussed how a conditional density/mass function represents the probabilities associated with one variable, conditional on the value of another.
- Since conditional PMFs/PDFs are still just mass/density functions, we can talk about the centre of these distributions.
- Therefore we can talk about a conditional expectation.
- Expectations based on conditional distributions are analogous to expectations based on one variable/marginal distributions.

Definitions of Conditional Expectations

- If the conditional mass function of $Y \mid X = x$ is discrete, then we find the conditional expectation as

$$E(Y \mid X = x) = \sum_y y p_{Y|X}(y \mid x)$$

  where recall that $p_{Y|X}(y \mid x)$ is a function of y.

- If the conditional density function is continuous, then we have

$$E(Y \mid X = x) = \int y f_{Y|X}(y \mid x) dy.$$

- More generally, this can be written for any function $h(Y)$ as

$$E\left(h(Y) \mid X = x\right) = \int h(y) f_{Y|X}(y \mid x) dy$$

- The discrete version is analogous.

Example: Gasoline exercise again

We previously saw that when we let X be the proportion of gasoline in the tank, and Y the proportion from the tank that has been sold, we have the joint PDF $f(x, y) = 3x, \text{ if } 0 \leq y \leq x \leq 1$. We further found that the conditional distribution of Y given X was

$$f_{Y|X}(y \mid x = 0.75) = \frac{1}{0.75}, 0 \leq y \leq 0.75$$

Find the average proportion of gas sold, given that the tank was only $3/4$ full.

- We just use the definition of conditional expectation to find this:
$$E(Y \mid X = 0.75) = \int_0^{0.75} \frac{y}{0.75} dy = \frac{3}{2}\left(\frac{3}{4}\right)^2 = \frac{27}{32}$$

Law of Total Expectation

- The next result is similar to the idea of the Law of Total Probability from Week 2.
- Recall we can calculate the probability of an event by conditioning on all the other events that could happen at the same time.
- The difference here is that we are not finding a probability, but an expected value.
- The Law of Total Expectation is useful when we don't know the marginal distribution of our random variable of interest, but still want the expectation of, say, Y.

Law of Total Expectation

Law of Total Expectation

$$E(Y) = E[E(Y \mid X)]$$

Proof:

$\square$

Example: Random Sums

The following sum $T = \sum_{i=1}^{N} X_i$ is called a random sum, because it is a sum of random variables, but also the upper bound of the sum is also random. N has finite expectation, and the X_i are random variables that are independent of N and have common mean.

- e.g. N could represent the number of jobs in a single-server queue and the X_i the service time of each job.
- It would be reasonable to want to know the average time T that would be required to serve all jobs in the queue.
- To do this, we need to condition on the number of jobs N and find the average time for those, then average over possible values of N.

Example: Random Sums (cont.)

- We can use the Law of Total Expectation to find $E(T)$ by

$$E(T) = E\left[E(T \mid N)\right]$$

- Here we just need to realize that if the X_i represent the time for each job, then if we condition on $N = n$, we know how many jobs and thus how many X_i we have, thus:

$$E(T \mid N = n) = \sum_{i=1}^{n} E(X_i) = nE(X)$$

  because the X_i have a common mean.

- Now, if we write this last statement as random variables: $E(T \mid N) = NE(X)$ and this we have

$$E(T) = E\left[E(T \mid N)\right] = E\left[NE(X)\right] = E(N)E(X)$$

Exercise - Give it a try!

Suppose that in a system, a component and a backup unit both have mean lifetimes of μ. If the component fails, the system switches to the backup unit, but there is a probability p that it will fail to make the switch. Let T be the total lifetime and $X = 1$ if the system switches to backup successfully and $X = 0$ if it does not. What is the expected total lifetime of the system?

Law of Total Variance

- We have a similar result to the law of total expectation, but this one instead pertains to the variance of a random variable Y.
- It follows the same logic as the total expectation one, except a bit less straightforward.

Law of Total Variance

$$Var(Y) = Var\left[E(Y \mid X)\right] + E\left[Var(Y \mid X)\right]$$

Law of Total Variance

Proof:

Example: Random Sums again

Suppose we are in the same context as before, but now we assume that the X_i are independent with a common mean $E(X)$ and same variance $Var(X)$, and that $Var(N) < \infty$.

- We can find $Var(T)$ using this new result,

$$Var(T) = E[Var(T \mid N)] + Var[E(T \mid N)]$$

- We found previously that $E(T \mid N) = NE(X)$, so we can now write

$$Var[E(T \mid N)] = Var[NE(X)] = [E(X)]^2 Var(N)$$

- Also, by the same logic as the earlier example, we have

$$Var(T \mid N = n) = Var\left(\sum_{i=1}^{n} X_i\right) = nVar(X)$$

Example: Random Sums again

- This last statement implies that we can generalize to any value $N = n$, and thus represent it with the random variable N:

$$Var(T \mid N) = NVar(X)$$

- Now just take the expectation of this last line:

$$E[Var(T \mid N)] = E[NVar(X)] = E(N)Var(X)$$

- Putting this all together, we find that

$$Var(T) = [E(X)]^2 Var(N) + E(N)Var(X)$$

- In context, this means that the variance of the total time is average job time squared times variability in number of jobs, combined with expected number of jobs times variability in time per job.

Outline

Application of Conditional Distributions

- Recall from our discussion of Bayes' rule (Week 2) that there are two philosophical branches of statistics:
  - Frequentist statistics: where we assume that we are trying to determine a value of a parameter in a model that is fixed but unknown
  - Bayesian statistics: where we try to determine a value of a parameter in a model that is itself random and has a distribution of likely values
- To make a guess about the value of a parameter, we use sample data
- When we perform Bayesian statistics, we use conditional probability to determine what is the probability of the value of a parameter given the chances of observing the data we saw

Application of Conditional Distributions

A coin has a certain probability of landing on heads when it is spun on its edge, but this probability is not necessarily 0.5. Suppose we spin the coin n times and we let X represent the number of times heads comes up. We want to know what is the success probability θ that a head comes up, given the number of heads we see.

- Here we have that θ is unknown and we want to know if we can make a guess as to its value.
- Since θ is unknown, we can think of it as a possible value from a random variable Θ, which represents success probabilities of getting heads.
- We can talk about the distribution of possible values for this success probability θ.
- One option: all values of θ are equally likely, so we might say that $\Theta \sim Uniform(0, 1)$ where $0 \leq \theta \leq 1$

Application of Conditional Distributions (cont.)

- The distribution for Θ is called a prior distribution, representing our belief about how likely the θ values are.
- If we did know θ (so conditional on its value), we can easily write the density for X, the number of heads we would see in n spins:

$$f_{X|\Theta}(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \; x = 0, 1, \ldots, n$$

- This conditional distribution is called the likelihood, representing how likely the number of heads $X = x$ will be given a certain value of success probability $\Theta = \theta$

Application of Conditional Distributions (cont.)

- When we run studies/do research, we collect data and try to use this to get an idea for the possible values of some parameter.
- In our case, we want to use the number of heads X to get an idea for the possible values of Θ.
- So we want to find $f_{\Theta|X}(\theta \mid x)$
- We can use the prior $f_{\Theta}(\theta)$ and the likelihood $f_{X|\Theta}(x \mid \theta)$ and conditional probability rules to do this.
- By multiplicative law, we get

$$f_{X,\Theta}(x,\theta) = f_{X|\Theta}(x \mid \theta) f_{\Theta}(\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

where $x = 0, 1, \ldots, n$ and $0 \leq \theta \leq 1$.

Application of Conditional Distributions (cont.)

- This joint density represents the joint chances of seeing a specific x and θ.
- Now we need to take $f_{X,\Theta}(x,\theta)$ and turn this into a conditional distribution $f_{\Theta|X}(\theta \mid x)$.
- Since $f_{\Theta|X}(\theta \mid x) = {}^{f_{X,\Theta}(x,\theta)}/_{f_X(x)}$, we need to find the marginal for X.
- Just integrate out θ from $f_{X,\Theta}(x,\theta)$:

$$f_X(x) = \int_0^1 \binom{n}{x} \theta^x (1-\theta)^{n-x} d\theta = \frac{1}{n+1}, \; x = 0, 1, \ldots, n$$

- The details of how we got this integral can be found in the textbook.

Application of Conditional Distributions (cont.)

- Now we have all the pieces to find $f_{\Theta|X}(\theta \mid x)$.
- Start from the definition:

$$f_{\Theta|X}(\theta \mid x) = \frac{f_{X,\Theta}(x,\theta)}{f_X(x)} = (n+1)\binom{n}{x}\theta^x(1-\theta)^{n-x}$$

- By writing the binomial coefficient in its factorial form, we can convert it into a collection of Gamma functions:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} = \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)}$$

- We get finally that

$$f_{\Theta|X}(\theta \mid x) = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^x(1-\theta)^{n-x}$$

Application of Conditional Distributions (cont.)

- $f_{\Theta|X}(\theta \mid x)$ is called the posterior distribution.
- This can be seen to be a $Beta(x+1, n-x+1)$
- It represents the probability of certain values of success probability given that I have collected information on the number of heads.
- We have taken our guess about the density of $f_{\Theta}(\theta)$ and updated it using data, and subsequently the chances of seeing those values of X to get $f_{\Theta|X}(\theta \mid x)$.

Application of Conditional Distributions (cont.)

- Suppose I spin a coin 20 times and get a head 13 times. I want to know what is the long-run probability of landing on heads for my coin.
- I think this success probability is distributed as a Uniform(0, 1).
- Then I combine this prior belief with my data of 13 heads in 20 spins.
- Based on my posterior, I know the probability, given my data, that the success probability takes a certain value will follow a $Beta(14, 8)$.
- Given my data, does my success probability have a high chance of being larger than 0.5?
- Based on my Beta posterior, I find

$$P(\Theta > 0.5 \mid X = 13) = \int_{0.5}^{1} f_{\Theta|X}(\theta \mid x) d\theta = 0.91$$