

# Mathematics Bootcamp

## Part III: Probability and Distribution Theory

Fan Bu<sup>1</sup>   Kyle Burris<sup>1</sup>

<sup>1</sup>Department of Statistical Science  
Duke University

Graduate Orientation, August 2018

# Outline

## Probability Theory

## Random Variables

- Distribution Functions of Random Variables

- Transformations of Random Variables

- A Gentle Introduction to Distribution Theory

## Multivariate Random Variables

## A Brief Overview of Bayesian Ideas

# Probability Theory

# Conditional Probability and Independence

Starting with something familiar. Consider two events  $A$  and  $B$  with the sample space  $\Omega$ .

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

Furthermore, consider the following notion of independence for the same two events.  $A$  and  $B$  are independent if:

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

# Conditional Probability and Independence - Continued

Conditional Probability for more than two events. Let  $A_1, A_2, \dots$  be a partition of the sample space and let  $B$  be any set, then for  $i = 1, 2, \dots$ :

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{\infty} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

We can similarly extend the definition of independence to cases with more than two events. A collection of events  $A_1, \dots, A_n$  are considered mutually independent if for any subcollection  $A_{i_1}, \dots, A_{i_K}$  we have that:

$$\mathbb{P}(\cap_{j=1}^K A_{i_j}) = \prod_{j=1}^K \mathbb{P}(A_{i_j})$$

# Conditional Probability - Example

In morse code, information is represented as dots and dashes.  
Assume the following:

$$\mathbb{P}(\text{dot sent}) = \frac{3}{7}; \quad \mathbb{P}(\text{dash sent}) = \frac{4}{7}$$

Furthermore, we also know that  $\mathbb{P}(\text{dot received} | \text{dot sent}) = \frac{7}{8}$ .  
Find  $\mathbb{P}(\text{dot sent} | \text{dot received})$ .

## Conditional Probability - Example Cont.

In order to use Bayes Rule, we first need  $\mathbb{P}(\text{dot received})$ .

$$\begin{aligned}\mathbb{P}(\text{dot received}) &= \mathbb{P}(\text{dot received} \cap \text{dot sent}) + \\ &\quad \mathbb{P}(\text{dot received} \cap \text{dash sent}) = \frac{7}{8} \frac{3}{7} \\ &\quad + \frac{1}{8} \frac{4}{7} = \frac{25}{26}\end{aligned}$$

Applying Bayes Rule:

$$\begin{aligned}\mathbb{P}(\text{dot sent} | \text{dot received}) &= \frac{\mathbb{P}(\text{dot sent} \cap \text{dot received})}{\mathbb{P}(\text{dot sent})} \\ &= \frac{\frac{7}{8} \frac{3}{7}}{\frac{25}{26}}\end{aligned}$$

## Conditional Probability - Exercise

In the population the probability of an infectious disease is  $\mathbb{P}(D) = 0.01$ . The probability of testing positive if the disease is present is  $\mathbb{P}(+|D) = 0.95$ . The probability of a negative test given the disease is not present is  $\mathbb{P}(-|ND) = 0.95$ . What is the probability of the disease being present if the test is positive i.e.  $\mathbb{P}(D|+)$ ?



## Conditional Probability - Exercise Cont.

First find the probability of a positive test:

$$\begin{aligned}\mathbb{P}(+) &= \mathbb{P}(+|D)P(D) + \mathbb{P}(+|ND)P(ND) = 0.01 \cdot 0.95 + 0.05 \cdot 0.99 \\ &= 0.059\end{aligned}$$

Next, we can invoke Bayes Rule:

$$\mathbb{P}(D|+) = \frac{\mathbb{P}(D \cap +)}{\mathbb{P}(+)} = \frac{0.01 \cdot 0.95}{0.059} \approx 0.161$$

# Independence - Example

Consider an experiment of tossing two dice. The sample space is therefore:

$$\Omega = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (2, 6), \dots, (6, 6)\}$$

Further, we define the events:

$$A = \{\text{doubles appear}\}$$

$$B = \{\text{the sum is between 7 and 10}\}$$

$$C = \{\text{the sum is 2 or 7 or 10}\}$$

Are the events  $A, B, C$  mutually independent?

## Independence - Example Cont.

Note that the following can be found by enumeration:

$$\mathbb{P}(A) = \frac{1}{6}; \quad \mathbb{P}(B) = \frac{1}{2}; \quad \mathbb{P}(C) = \frac{1}{3}$$

Furthermore:

$$\begin{aligned}\mathbb{P}(A \cap B \cap C) &= \mathbb{P}(\text{sum is 8, comprised of doubles}) = \frac{1}{36} \\ &= \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{3}\end{aligned}$$

But notice that  $\mathbb{P}(B \cap C) = \frac{11}{36} \neq \mathbb{P}(B)\mathbb{P}(C)$ . Therefore we do not have pairwise independence and hence claims of mutual independence cannot be made.

## Independence - Exercise

Consider the following sample space that consists of the  $3!$  permutations of  $\{a, b, c\}$  along with triples of each letter:

$$\Omega = \{aaa, bbb, ccc, abc, bca, cba, acb, bac, cab\}$$

Each element in  $\Omega$  is assumed to have probability  $\frac{1}{9}$ . Define the event  $A_i$ :

$$A_i = \{i^{th} \text{ place in the triple is occupied by } a\};$$
$$i = 1, 2, 3$$

$$\mathbb{P}(A_i) = \frac{1}{3}$$

Are the events  $A_i$  mutually independent?

## Independence - Exercise Cont.

Pairwise independence is satisfied:

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_2 \cap A_3) = \frac{1}{9}$$

But the joint event:

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{9} \neq \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$$

Hence, the events are **not** mutually independent

# Random Variables

# Random Variables

## Definition:

A *random variable* is a function from the sample space to the real numbers

**Note:** For those of you taking STA-711 you will learn a more formal definition

# Random Variables - Example

**The Experiment:** 2 Dice are rolled together

**The Sample Space:** All pairs of numbers from 1 through 6

**The Random Variable:** The sum of the numbers



# Random Variables - Exercise

**The Experiment:** A coin is tossed 5 times

**The Sample Space:**  $2^5$  possible permutations

**The Random Variable:**

**Note:** There is more than one right answer here

# Cumulative Distribution Functions of Random Variables

**Definition:** The cumulative distribution function (CDF) of a random variable denoted by  $F_X(x)$  is defined as:

$$F_X(x) = P_X(X \leq x); \quad \forall x$$

A function is a CDF if and only if the following are true:

- ▶  $\lim_{x \rightarrow -\infty} F(x) = 0$  and  $\lim_{x \rightarrow \infty} F(x) = 1$
- ▶  $F(x)$  is a non-decreasing function of  $x$
- ▶  $F(x)$  is right continuous i.e. for every number  $x_0$ ,  $\lim_{x \rightarrow x_0} F(x) = F(x_0)$

An important implication of CDFs: *A random variable  $X$  is continuous if  $F_X(x)$  is a continuous function of  $x$ . A random variable is discrete if  $F_X(x)$  is a step function of  $x$ .*

# Cumulative Distribution Functions of Random Variables - Example

If  $p$  denotes the probability of getting a head on any toss, and the experiment consists of tossing a coin until a head appears, then we define the random variable  $X$  = the number of tosses required until a head. The CDF of this random variable is given as:

$$P(X \leq x) = \sum_{i=1}^x (1-p)^{i-1} p$$

# Density and Mass Functions of Random Variables

Related to any random variable  $X$  and its CDF are the concept of probability *density* and probability *mass* functions. Specifically, a *probability mass function* (PMF) for a discrete random variable is defined as:

$$f_X(x) = P(X = x); \forall x$$

and the *probability density function* (PDF) for a continuous random variable is defined as a function that satisfies the following relationship:

$$F_X(x) = \int_{-\infty}^x f_X(t) dt; \forall x$$

# Density and Mass Functions of Random Variables - Example

An example of a density function for a Geometric Random variable from the coin tossing example earlier:

$$f_X(x) = P(X = x) = (1 - p)^{x-1} p \cdot \mathbf{1}(x \in 1, 2, 3, \dots)$$

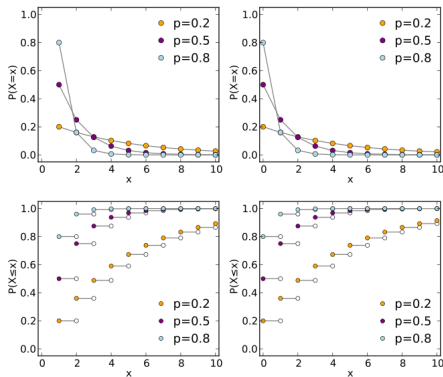
Notice that we can use the PMF (and analogously the PDF) to derive the CDF:

$$P(X \leq b) = \sum_{k=1}^b f_X(k) = F_X(b)$$

This partial sum is what we had used to reach the geometric CDF presented earlier

# Density and Mass of Random Variables - Example Cont.

Consider the following illustrations, courtesy of Wikipedia:



**Figure: Top Panel:** The PMF of the geometric distribution under both parameterizations **Bottom Panel:** The CDF of the geometric distribution under both parameterizations

# Transformations of Random Variables using the Change of Variables Formula

Assume that  $X$  has a pdf  $f_X(x)$  and that  $Y = g(X)$  where  $g$  is a monotone function. Suppose that  $f_X(x)$  is continuous on  $\mathcal{X}$ , and that  $g^{-1}$  has a continuous derivative on  $\mathcal{Y}$  where  $\mathcal{X}, \mathcal{Y}$  are such that  $\mathcal{X} = \{x : f_X(x) > 0\}$  and  $\mathcal{Y} = \{y : y = g(x)\}$ . Then the pdf of  $Y$  is given as follows:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

# Transformations of Random Variables - Example

Assume that  $X \sim f_X(x) = 1$  i.e.  $X \sim \text{Uniform}(0, 1)$ . Furthermore,  $Y = -\log(X)$ . What is the PDF of  $Y$ ?

First note that  $g(X) = Y = -\log(X) \rightarrow g^{-1}(Y) = e^{-Y}$ .

Therefore, using the formulation from earlier:

$$f_Y(y) = 1 \cdot |-e^{-Y}| = e^{-Y}$$

$$Y \sim \text{Exponential}(\lambda = 1)$$



# Transformations of Random Variables - Exercise

Assume that  $X \sim \text{Normal}(0, 1)$ . Let  $Y = X^2$ . What is the distribution of  $Y$ ?

The PDF of the standard normal distribution is given as follows:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$$

## Transformations of Random Variables - Exercise Cont.

Consider that  $Y = g(X) = X^2 \rightarrow g^{-1}(Y) = \mp\sqrt{Y}$ . Hence, consider that we can partition the support of  $X$  into two pieces  $S_1 = (-\infty, 0)$  and  $S_2 = (0, \infty)$  where the function  $g(X)$  is monotone. Note that  $\mathcal{Y} = (0, \infty)$ . Use the change of variables formulation over the two partitions and sum:

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(-\sqrt{Y})^2}{2}\right\} \left| -\frac{1}{2\sqrt{y}} \right| + \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(\sqrt{Y})^2}{2}\right\} \left| \frac{1}{2\sqrt{y}} \right| \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{Y}} \exp\left\{-\frac{y}{2}\right\} \end{aligned}$$

Hence, we get that  $Y \sim \chi^2_{df=1}$

# Expectations and Variances of Random Variables

The expectation of any random variable can be computed as follows:

- ▶  $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$  when  $X$  is continuous
- ▶  $\mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x)f_X(x) = \sum_{x \in \mathcal{X}} g(x)\mathbb{P}(X = x)$  when  $X$  is discrete

The variance can be computed using the expectations as follows:

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

You will need to do some calculus to find each of these quantities

# Kernel Tricks for Computing Expectations - Example

If we say that  $X \sim \text{Exponential}(\lambda)$ , with PDF  $f_X(x) = \lambda \exp\{-\lambda x\}$ . In order to find  $\mathbb{E}[X]$ , you must find:

$$\mathbb{E}[X] = \int_0^{\infty} x \lambda \exp\{-\lambda x\} dx$$

- ▶ Integration by parts
- ▶ Something a bit more clever?

## Kernel Tricks for Computing Expectations - Example Cont.

First, notice that if we say that  $X \sim \text{Gamma}(\alpha, \beta)$  and PDF  $g_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$

In instances when  $\alpha = 1$  then this an Exponential random variable with  $\lambda = \beta^1$ .

The integrand from the previous slide, is *almost* like a Gamma PDF with  $\alpha = 2$ . Hence, you can complete it by some clever multiplication and division:

$$\mathbb{E}[X] = \frac{\Gamma(2)}{\lambda^2} \int_0^\infty \frac{\lambda^2}{\Gamma(2)} x^{2-1} \lambda \exp\{-\lambda x\} dx = \frac{1}{\lambda} \cdot 1 = \frac{1}{\lambda}$$

---

<sup>1</sup>Note that is using the rate parameterization. Sometimes when we use the scale parameterization the reciprocal of the  $\lambda$  parameter is used

# Kernel Tricks for Computing Expectations - Exercise

Use the kernel trick for Exponential random variables to find  $\mathbb{V}[X]$

# Kernel Tricks for Computing Expectations - Exercise

The first step in this process is to find  $\mathbb{E}[X^2]$  which you can calculate using the kernel trick as follows:

$$\mathbb{E}[X^2] = \frac{\Gamma(3)}{\lambda^3} \int_0^\infty \frac{\lambda^3}{\Gamma(3)} x^{2-1} \lambda \exp\{-\lambda x\} dx = \frac{2}{\lambda^2} \cdot 1 = \frac{2}{\lambda^2}$$

Plug this result into the variance formula presented earlier using the expectation from earlier:

$$\mathbb{V}[X] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

**Note:** You will not have to use a lot of integration here. Always try these tricks first

# Properties of Expectations and Variances

Assume a random variable  $X$  and  $a$  is a scalar constant, then:

$$\mathbb{E}[aX] = a\mathbb{E}[X]$$

$$\mathbb{V}[aX] = a^2\mathbb{V}[X]$$

Variances also have nice properties. Consider two random variables  $X$  and  $Y$ .

$$\mathbb{V}[X \mp Y] = \mathbb{V}[X] + \mathbb{V}[Y] \mp 2\mathbb{C}(X, Y)$$

These extend to multivariate random variables as well. We will see these in the second session today



# Discrete Distributions

A random variable  $X$  is *discrete* if the range of  $X$ , the sample space, is countable. In most situations, the random variable has integer valued outcomes

Some example of discrete distributions:

- ▶ Binomial Distribution
- ▶ Poisson Distribution

# Binomial Distribution

This distribution counts the the number of successes in  $n$  independent trials all with the same fixed probability  $p$  of success

$$X \sim \text{Binomial}(n, p)$$

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$\mathbb{E}[X] = np$$

$$\mathbb{V}[X] = np(1-p)$$

# Poisson Distribution

This distribution is used for counting the number of events over some time horizon based on an intensity parameter  $\lambda$

$$X \sim \text{Poisson}(\lambda)$$

$$P(X = x) = \frac{\exp^{-\lambda} \lambda^x}{x!}$$

$$\mathbb{E}[X] = \mathbb{V}[X] = \lambda$$

## Poisson Distribution - Exercise

Prove that  $\mathbb{E}[X] = \lambda$  if  $X \sim \text{Poisson}(\lambda)$

## Poisson Distribution - Exercise Cont.

We need to compute the following:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x=0}^{\infty} x \frac{\exp^{-\lambda} \lambda^x}{x!} \\ &= \lambda \exp\{-\lambda\} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}\end{aligned}$$

Now recognize the following result from the Taylor series expansion on  $\exp\{y\} = \sum_{i=0}^{\infty} \frac{y^i}{i!}$ . Use this result with a clever substitution:

$$\lambda \exp\{-\lambda\} \exp\{\lambda\} = \lambda$$

# Continuous Distributions

A random variable  $X$  is *continuous* if the range of  $X$ , the sample space, takes on an uncountably infinite number of values. In most instances the random variable has real-valued outcomes.

Some examples of Continuous Distributions

- ▶ Gamma Distribution
- ▶ Inverse-Gamma Distribution
- ▶ Normal Distribution

# Gamma Distribution

A random variable variable  $X \sim \text{Gamma}(\alpha, \beta)$  with PDF:

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-x\beta\}$$

$$\mathbb{E}[X] = \frac{\alpha}{\beta}$$

$$\mathbb{V}[X] = \frac{\alpha}{\beta^2}$$

$$\alpha, \beta > 0$$

$$x \in (0, \infty)$$

# Gamma Distribution - Alternative Parameterization

We often say that  $X \sim \text{Gamma}(k, \theta)$  with PDF:

$$f_X(x) = \frac{\theta^{-k}}{\Gamma(k)} x^{k-1} \exp\left\{-\frac{x}{\theta}\right\}$$

$$\mathbb{E}[X] = k\theta$$

$$\mathbb{V}[X] = k\theta^2$$

$$k, \theta > 0$$

$$x \in (0, \infty)$$

Note that in this framework  $\alpha = k$  and  $\theta = \frac{1}{\beta}$



# Gamma Distribution - Important Properties

Here are some important tricks that will be useful in **711** and **601**

- ▶ if  $\alpha = 1$  and then  $X \sim \text{Exponential}(\lambda = \beta)$
- ▶ if  $\alpha = \frac{\nu}{2}$  and  $\beta = \frac{1}{2}$  then  $X \sim \chi^2_\nu$
- ▶ if  $X \sim \text{Gamma}(\alpha_1, \beta)$  and  $Y \sim \text{Gamma}(\alpha_2, \beta)$  then  
 $X + Y \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$
- ▶ if  $X \sim \text{Gamma}(k, \theta)$ , then  $\frac{1}{X} \sim \text{Inverse - Gamma}(k, \frac{1}{\theta})$

# Normal Distribution

A random variable  $X \sim \text{Normal}(\mu, \sigma^2)$  with PDF:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}$$

We also sometimes express this in terms of a *precision* parameter, rather than a variance,  $X \sim \text{Normal}(\mu, \phi^{-1})$  which will become useful ... everywhere since we are from the cult of Bayes

# Exponential Families

A family of PDFs and PMFs are called exponential family distributions if they can be expressed in the form:

$$f_X(x|\theta) = h(x)c(\theta) \exp\left\{\sum_{i=1}^k \omega_i(\theta)t_i(x)\right\}$$

Where:

$$h(x) \geq 0$$

$$c(\theta) = 0$$

$$\omega_1(x), \dots, \omega_k(x) \in \mathbb{R}$$

$$t_1(x), \dots, t_k(x) \in \mathbb{R}$$

While this may not seem important yet, exponential family distributions have some really, really nice properties that make their applications extremely widespread (**732**)

## Exponential Families - Example

Consider the binomial PMF for a random variable  $X \sim \text{Binomial}(n, p)$

$$P(X = x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}$$

This is an exponential family PMF. We can show this by re-expressing terms:

$$P(X = x) = \frac{n!}{(n-x)!x!} (1-p)^n \exp\left\{x \log\left(\frac{p}{1-p}\right)\right\}$$

$$h(x) = \frac{n!}{(n-x)!x!} \mathbb{I}_{x=0,\dots,n}$$

$$c(p) = (1-p)^n$$

$$\omega_1(p) = \log\left(\frac{p}{1-p}\right)$$

$$t_1(x) = x$$

# Exponential Families - Exercise

Consider the following normal PDF for  $X \sim \text{Normal}(\mu, \sigma^2)$

$$f_X(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}$$

Show that this is an exponential family PDF

## Exponential Families - Exercise Cont.

Consider the following PDF

$$f_X(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}$$

Expanding the exponential yields the following

$$f_X(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\mu^2}{2\sigma^2}\right\} \exp\left\{-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\right\}$$

$$h(x) = 1$$

$$c(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-\mu^2}{\sigma^2}\right\}; \mu \in \mathbb{R} \quad \sigma^2 > 0$$

$$\omega_1(\mu, \sigma) = \frac{1}{\sigma^2} \quad \omega_2 = \frac{\mu}{\sigma^2}$$

$$t_1(x) = \frac{-x^2}{2} \quad t_2(x) = x$$

# Multivariate Random Variables

## Random Vectors

The basic definition of a *Random Vector* carries over from the definition of a random variable presented earlier. In the later half, you'll see this is a more general setting. Here we will emphasize bi-variate examples.

An  $n$  dimensional random vector is a function from a sample space into  $\mathbb{R}^n$ ,  $n$ -dimensional euclidean space

An example of a 2-dimensional random vector is as follows:

Recall that two dice being rolled have a sample space of 36 points. Associate the random variables  $X$  and  $Y$  with the sample space as follows:

$$X = D_1 + D_2$$

$$Y = |D_1 - D_2|$$

In this way, the pair  $(X, Y)$  define a bi-variate random vector



# Distribution Functions for Multivariate Random Variables

There are three types of distribution functions that we will cover:

- ▶ Joint Distribution
- ▶ Marginal Distribution
- ▶ Conditional Distribution

# Joint Distribution - Bivariate Case

**Joint PDF:** A function  $f(x, y)$  from  $\mathbb{R}^2 \rightarrow \mathbb{R}$  is called a joint PDF of the random vector  $(X, Y)$  if for every  $A \subset \mathbb{R}^2$

$$\mathbb{P}((X, Y) \in A) = \int_A \int f_{X,Y}(x, y) dx dy$$

**Joint PMF:** The function  $f(x, y)$  from  $\mathbb{R}^2 \rightarrow \mathbb{R}$  defined by  $f_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$  is the joint PMF of  $X, Y$ . Then for every  $A \subset \mathbb{R}^2$

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

## Joint Distribution - Exercise

Assume that  $X$  and  $Y$  have the joint PDF:

$$f_{X,Y}(x,y) = 4xy$$

$$0 < x < 1$$

$$0 < y < 1$$

Find  $P(Y < X)$

## Joint Distribution - Exercise Cont.

We can set up the double integral required for this probability as follows:

$$\begin{aligned} p(Y < X) &= \int_0^1 \int_0^x 4xy \, dy \, dx \\ &= \int_0^1 \left[ 4x \frac{y^2}{2} \right]_0^x \, dx \\ &= \int_0^1 2x^3 \, dx = \frac{1}{2} \end{aligned}$$

# Marginal Distribution

Given the joint PDF or joint PMF, you can find the marginal PDF or PMF:

**Marginal PDF:**

$$f_X(x) = \int_Y f_{X,Y}(x,y) dy$$

**Marginal PMF:**

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

# Conditional Distribution

Assume that  $X, Y \sim f_{X,Y}(x, y)$ , then we can employ Bayes' rule for distributions:

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

# Multivariate Distributions - Exercise

Assume that  $(X, Y)$  are a continuous random vector with joint pdf given by:

$$f_{X,Y}(x, y) = \exp\{-y\} \quad 0 < x < y < \infty$$

Find the marginal distribution of  $X$  and the conditional distribution  $Y|X$

## Multivariate Distributions - Example Cont.

We start by finding the marginal distribution of  $X$ :

$$f_X(x) = \int_x^\infty \exp\{-y\} dy = e^{-x}$$
$$X \sim \text{Exponential}(\lambda = 1)$$

Now use the results on conditional distributions given earlier, to find:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\exp\{-y\}}{\exp\{-x\}} \mathbb{I}(x < y)$$



# Total Expectation and Total Variance Laws

In many examples, you are interested in marginal moments from conditional distributions. Your first option of course is to find the joint distribution, do some marginalization and then integrate, but I do not like calculus so **instead**:

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

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

# Total Expectation and Total Variance Laws - Example

Assume that we have the following relationship:

$$X|N \sim \text{Binomial}(N, p)$$

$$N \sim \text{NegativeBinomial}(\tau = \frac{1}{1+\beta}, r = 1)$$

Find  $\mathbb{E}[X]$  and  $\mathbb{V}[X]$

**Tip:** that  $\mathbb{E}[N] = \frac{r\tau}{1-\tau}$  and  $\mathbb{V}[N] = \frac{\tau r}{(1-\tau)^2}$

# Total Expectation and Total Variance Laws - Example Cont.

First, we iterate to find the expectation

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[\mathbb{E}[X|N]] \\ &= \mathbb{E}[Np] \\ &= p \frac{1}{1+\beta} \\ &= \frac{p}{\beta}\end{aligned}$$

Next, we proceed with finding the variance

$$\begin{aligned}\mathbb{V}[X] &= \mathbb{E}[\mathbb{V}[X|N]] + \mathbb{V}[\mathbb{E}[X|N]] \\ &= \mathbb{E}[Np(1-p)] + \mathbb{V}[Np] \\ &= \frac{p(1-p)}{\beta} + p^2 \frac{1+\beta}{\beta^2}\end{aligned}$$

# Total Expectation and Total Variance Laws - Exercise

$$\begin{aligned}X|P &\sim \text{Binomial}(n, P) \\ P &\sim \text{Beta}(a, b)\end{aligned}$$

Find the  $\mathbb{E}[X]$  and  $\mathbb{V}[X]$

**Tip:**

$$\begin{aligned}\mathbb{E}[P] &= \frac{a}{a+b} \\ \mathbb{V}[P] &= \frac{ab}{(a+b)^2(a+b+1)}\end{aligned}$$

# Total Expectation and Total Variance Laws - Exercise Cont.

We can start by finding the marginal expectation first:

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|P]] = \mathbb{E}[nP] = n\mathbb{E}[P] = n\frac{a}{a+b}$$

And then the marginal variance:

$$\begin{aligned}\mathbb{V}[X] &= \mathbb{V}[\mathbb{E}[X|P]] + \mathbb{E}[\mathbb{V}[X|P]] \\ &= \mathbb{V}[nP] + \mathbb{E}[nP(1-P)] \\ &= n^2\mathbb{V}[P] + n\mathbb{E}[P - P^2] \\ &= n^2\frac{ab}{(a+b)^2(a+b+1)} + n\frac{a}{a+b} \\ &\quad - n\left(\frac{ab}{(a+b)^2(a+b+1)}\right) - n\left(\frac{a}{a+b}\right)^2 \\ &= n\frac{ab(a+b+n)}{(a+b)^2(a+b+1)}\end{aligned}$$

# A Brief Overview of Bayesian Ideas

# Bayes Rule

You have already seen Bayes' Rule for events:

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$$

and also in the context of distributions:

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

But what if  $Y$  was some **data** you had observed, and  $X$  was instead a parameter that controlled the distribution of  $Y$ , and was itself a random variable that you wanted to learn something about?

# The Bayesian Framework

We'll now use a slightly altered version of Bayes' rule. Assume that we observe the data  $Y$  and it is theorized to follow some distribution which has a parameter  $\theta$  that you want to conduct some inference on.

$$f(\theta|Y) \propto f(\theta)f(Y|\theta)$$

Where:

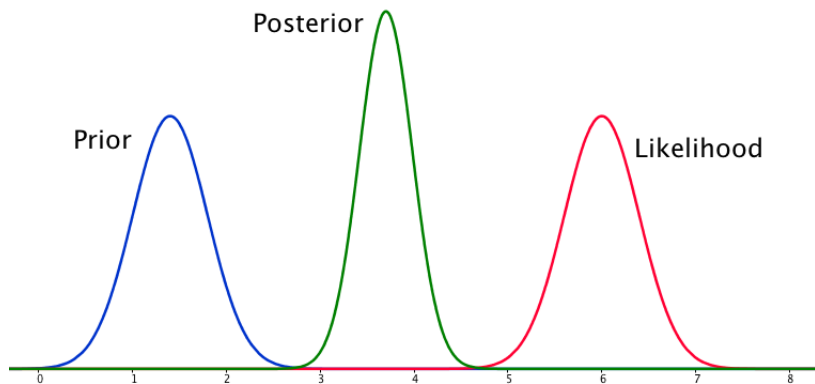
- ▶  $f(\theta|Y)$ : The posterior
- ▶  $f(\theta)$ : The prior
- ▶  $f(Y|\theta)$ : The likelihood

The *posterior* distribution allows you to conduct inference on the parameter  $\theta$  in light of the data that you observe via the *likelihood*, regularized by beliefs quantified in the *prior*. The easiest method of inference is via *conjugacy*.



# The Bayesian Framework Cont.

Relationship between the posterior, likelihood and prior:



**Figure:** The posterior can be thought of as a compromise between the data and expertise quantified in the prior

# Conjugacy of the Prior Distribution

**Definition:** A prior distribution is considered conjugate for a likelihood if the resulting posterior distribution is in the same family as the prior.

The example that follows will attempt to make this idea more concrete.

# Conjugacy of the Prior - Example

Assume that you observe  $n$  observations theorized to follow a Poisson distribution  $Y_1, \dots, Y_n \sim \text{Poisson}(\lambda)$ . We can further assume that  $\lambda \sim \text{Gamma}(a, b)$ . Find the posterior distribution of the parameter  $\lambda | Y_1, \dots, Y_n$ .

## Conjugacy of the Prior - Example Cont.

The Poisson likelihood is given as follows:

$$\begin{aligned}f(y_1, \dots, y_n | \lambda) &= \prod_{i=1}^n \exp\{-\lambda\} \frac{1}{y_i!} \lambda^{y_i} \\ &\propto \exp\{-n\lambda\} \lambda^{n\bar{y}}\end{aligned}$$

The Gamma prior is given as follows:

$$\begin{aligned}f(\lambda) &= \frac{b^a}{\Gamma a} \lambda^{a-1} \exp\{-\lambda b\} \\ &\propto \lambda^{a-1} \exp\{-\lambda b\}\end{aligned}$$

Using Bayes' rule we can find the posterior as follows:

$$\begin{aligned}f(\lambda | y_1, \dots, y_n) &\propto f(y_1, \dots, y_n | \lambda) f(\lambda) \\ &\propto \exp\{-n\lambda\} \lambda^{n\bar{y}} \theta^{a-1} \exp -\theta b \\ &\propto \exp\{-(b + n)\} \lambda^{n\bar{y} + a - 1}\end{aligned}$$

This is exactly the kernel of the Gamma distribution:

$$\lambda | y_1, \dots, y_n \sim \text{Gamma}(n\bar{y} + a, b + n)$$

## Conjugacy of the Prior - Exercise

Assume that you observe  $n$  observations theorized to follow a bernoulli distribution  $Y_1, \dots, Y_n \sim \text{Bernoulli}(p)$ . We can further assume that  $p \sim \text{beta}(a, b)$ . Find the posterior distribution of the parameter  $p|Y_1, \dots, Y_n$ .

## Conjugacy of the Prior - Exercise Cont.

The Bernoulli likelihood is given as follows:

$$\begin{aligned}f(y_1, \dots, y_n | p) &= \prod_{i=1}^n (1-p)^{1-y_i} p^{y_i} \\ &\propto p^{n\bar{y}} (1-p)^{n-n\bar{y}}\end{aligned}$$

The prior for the parameter  $p$  is given  $\text{Beta}(a, b)$

$$\begin{aligned}f(p) &= \frac{1}{\beta(a, b)} p^{a-1} (1-p)^{b-1} \\ &\propto p^{a-1} (1-p)^{b-1}\end{aligned}$$

Using Bayes' rule again yields:

$$\begin{aligned}f(p | y_1, \dots, y_n) &\propto f(y_1, \dots, y_n | p) f(p) \\ &\propto p^{n\bar{y}} (1-p)^{n-n\bar{y}} p^{a-1} (1-p)^{b-1} \\ &\propto p^{n\bar{y}+a-1} (1-p)^{n+b-n\bar{y}-1}\end{aligned}$$

This is exactly the kernel of the Beta distribution:

$$p | y_1, \dots, y_n \sim \text{Beta}(a + n\bar{y}, b + n(1 - \bar{y}))$$

# Questions

# Reference Guide

- ▶ *Statistical Inference* - Casella and Berger
- ▶ *Mathematical Statistics* - Bickel and Doksum
- ▶ *A First Course in Bayesian Statistical Methods* - Hoff
- ▶ *Bayesian Computation with R* - Albert