Contents

Chapter 1

Population Proportion Statistics

Let's say that you are trying to get a candidate elected. The candidate asks you "What proportion of the voting population is going to vote for me?" So you go out and ask a random sample of 12 voters. 11 say that they are going to vote for your candidate. What can you tell the candidate?

1.1 Sample Probabilities from Population Proportion

To address these sorts of questions (and there are a lot of them), we start with the opposite question: If we knew what the proportion was in the entire population, what sort of results should we expect in a random sample of just 12?

For example, lets say that 62% of the entire population plan to vote for your candidate. You ask 12 "Will you vote for my candidate?" How many will say "Yes"? You don't know – it depends on the sample. For example, there is some chance that you will just happen to choose all 12 from the 48% of the population that doesn't plan to vote for your candidate.

We can compute the probability of each outcome using the binomial distribution. Let r be the probability that a random person will say "I plan to vote for your candidate." Let n be the number of people you ask. The probability that exactly k people will say "I plan to vote for your candidate" is given by:

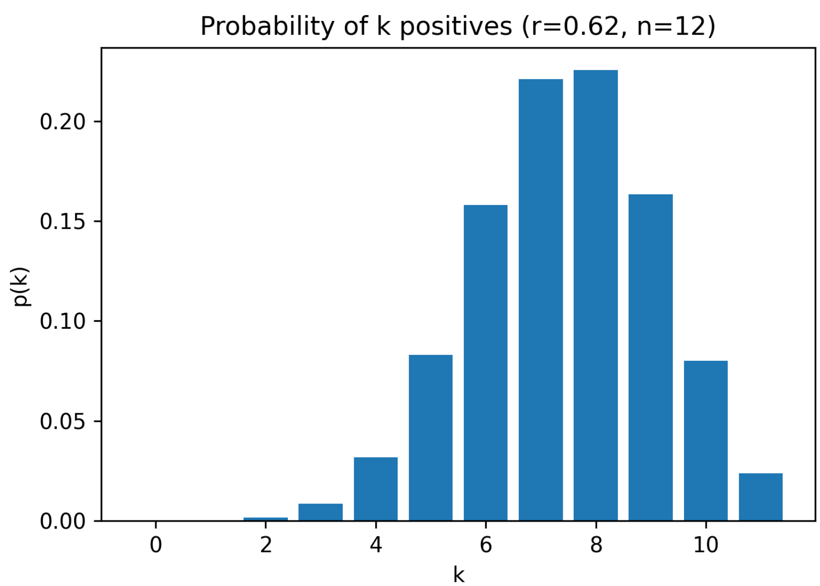
$$p(k) = \binom{n}{k} r^k (1-r)^{n-k}$$

Note that even though most people support your candidate, there is some chance that no one you ask will say that they will vote for your candidate.

Using $r = 0.62$, we can compute the probability of each outcome:

k	p(k)
0	0.000009
1	0.000177
2	0.001593
3	0.008663
4	0.031801
5	0.083017
6	0.158024
7	0.220996
8	0.225358
9	0.163418
10	0.079989
11	0.023729

Looking at this, the most likely outcome is that 8 people will say "Yes." However, there is less than a 1 in 4 chance of that outcome. It is very unlikely that less than 2 people will say "Yes." Here is a bar chart of the data



In this section, we knew the proportion of the population (p) and used that to find the probability of each possible number of positives in a random sample (k). Now we are going to go the other way: You know k, and you are finding the probability of possible values of p.

1.2 Population Proportion from Sample

You ask 12 people if they will vote for your candidate. 9 say "Yes."

Now you do a thought experiment: "If only 10% of the population were going to vote for my candidate, what is the probability that I would see this outcome?"

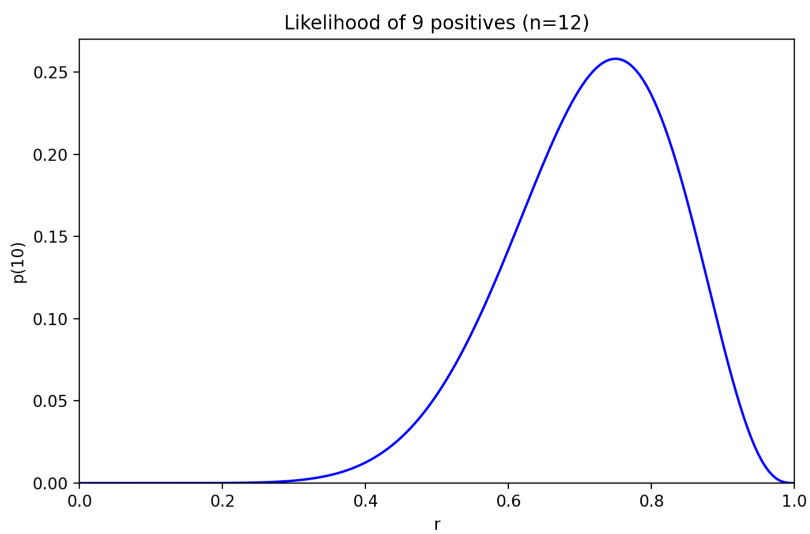
$p(9) = \binom{n}{k} r^k (1-r)^{n-k} = \binom{12}{9} (0.1)^9 (0.9)^{12-9} = 0.00000016$

So this outcome would be quite unusual. What if 70% of the population were going to vote for your candidate? What is the probability that you would see this outcome?

$p(9) = \binom{n}{k} r^k (1-r)^{n-k} = \binom{12}{9} (0.7)^9 (0.3)^{12-9} \approx 0.2397$

In this case, the observed outcome would be a lot less unusual.

So you decide to plot out the likelihood of this outcome for every possible value of r:



This looks a lot like a probability distribution, but *it is not*. The area under the curve does not integrate to 1.0 – it is significantly less. This is a called a *likelihood*.

However, it still tells us something, right? The maximum likelihood estimator is $9/12 = 0.75$.

1.3 From Likelihood to Probability Density Function

How can we make this likelihood into a probability density function? We use Bayes' Law for continuous probability:

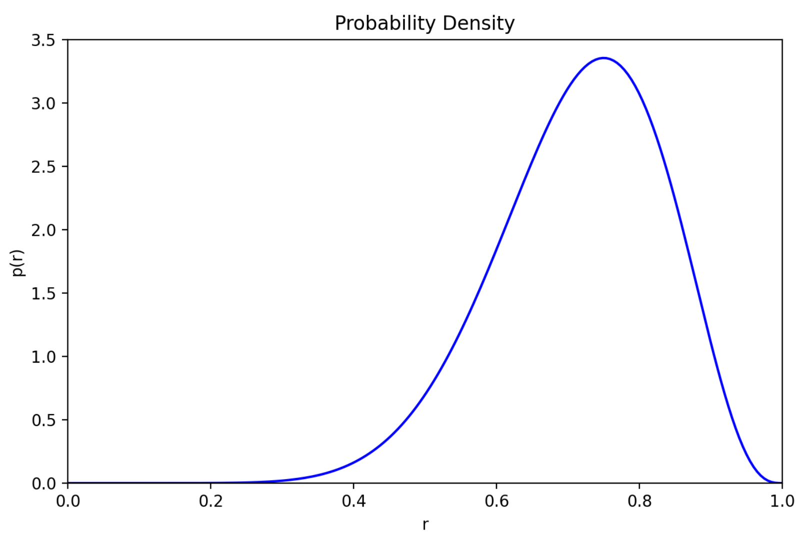
$$p(r|k) = \frac{p(k|r)p(r)}{\int_{r=0}^{1} p(k|r)p(r)dr}$$

That is, given that we had k positive responses, what is the probability that the proportion of the population that will vote for your candidate is r? The numerator of the fraction is the likelihood scaled up or down by our prior belief about the value of r. What is the denominator of the fraction? For this to be a probability distribution, we need it to integrate to 1. This is taken care of by the denominator.

Let's say we have no prior belief about the value of r. That is $p(r)$ is the continuous uniform distribution between 0 and 1, thus $p(r) = 1$ for all possible values of r. Our formula becomes:

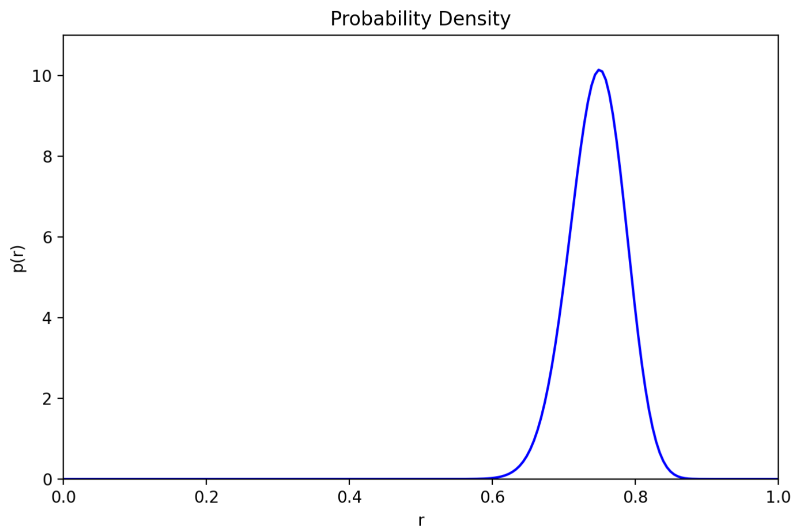
$$p(r|k) = \frac{p(k|r)}{\int_{r=0}^{1} p(k|r)dr}$$

That is the likelihood scaled up so that it integrates to 1. If we plot this, we get:



Here then, is your report to your candidate: "I asked 12 voters if they were going to vote for you. 9 said yes. Using a uniform prior, here is what I believe about your support in the general population." And include this graph.

What happens to this graph if you ask 120 voters and 90 say yes?



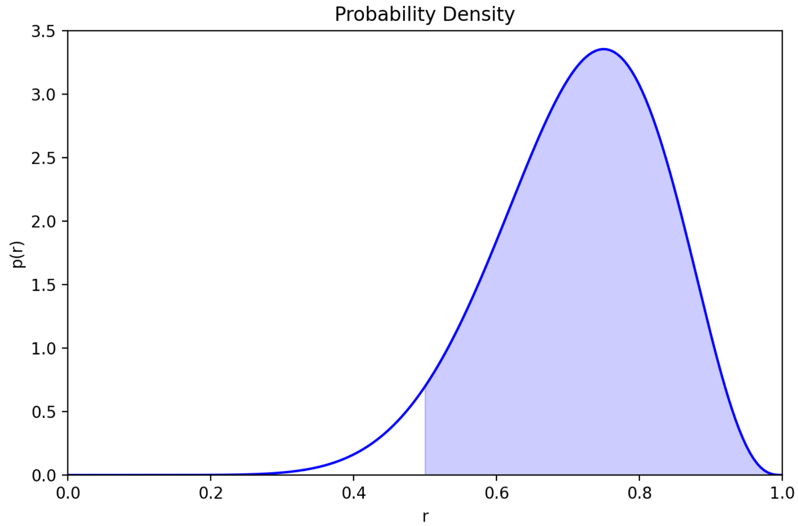
The MLE (0.75) is the same, but because of the much larger sample size, you are more confident when you say ”It is probably close to 75%.”

1.4 Beta Distribution

This probability distribution that you discovered is actually pretty common. It is known as the *beta distribution*.

The beta distribution has two parameters a and b that determine its shape. If you get k positives out of n, then use $a = k + 1$ and $b = n - k + 1$.

When you make your report to your candidate, they will look at your probability distribution with quiet awe and ask ”Based on your sample of 12 people, what is the probability that at least 50% of the population will vote for me?” So, you’d fill in the region and say ”This area represents that probability.”



Once again, there will be a long silence. And then they will ask "Can you give me a number?" Here is the python code:

```
import numpy as np
from scipy.stats import beta

# Constants
K =  9
N = 12

# What is the probability r <=0.5?
p_less = beta.cdf(0.5, K +1, N - K +1)

# What is the probability r > 0.5?
p_more = 1.0 - p_less
print(f"I'm {p_more * 100.0:.2f}% sure you will win.")
```

This will give you:

```
I'm 95.39% sure you will win.
```

Chapter 2

The Normal Distribution

The Normal distribution, also known as the Gaussian distribution, is a type of continuous probability distribution for a real-valued random variable. It is one of the most important probability distributions in statistics due to its several unique properties and usefulness in many areas.

2.1 Defining the Normal Distribution

The Normal distribution is defined by its mean (μ) and standard deviation (σ). The probability density function (pdf) of a Normal distribution is given by:

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

where:

- x is the point up to which the function is integrated,
- μ is the mean or expectation of the distribution,
- σ is the standard deviation,
- σ^2 is the variance.

2.2 Importance of the Normal Distribution

There are several reasons why the Normal distribution is crucial in statistics:

- **Central Limit Theorem:** One of the main reasons for the importance of the Normal distribution is the Central Limit Theorem (CLT). The CLT states that the distribution of the sum (or average) of a large number of independent, identically distributed variables approaches a Normal distribution, regardless of the shape of the original distribution.
- **Symmetry:** The Normal distribution is symmetric, which simplifies both the theoretical analysis and the interpretation of statistical results.

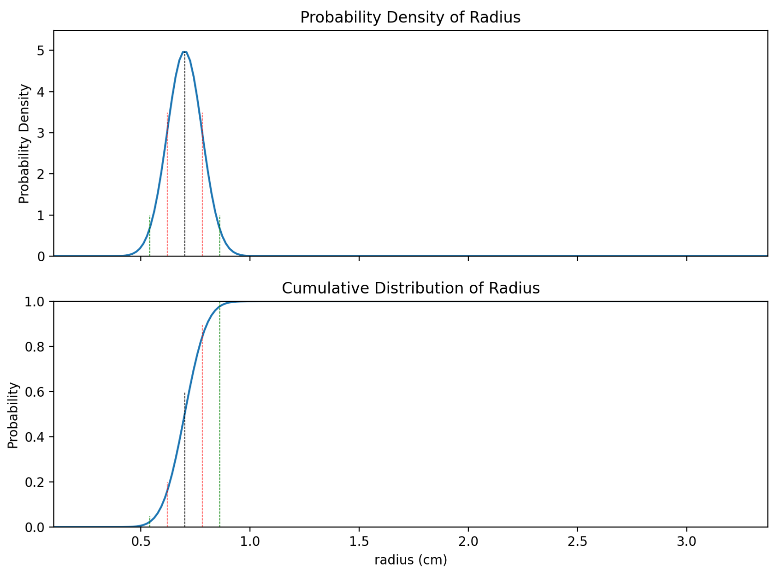
- **Characterized by Two Parameters:** The Normal distribution is fully characterized by its mean and standard deviation. The mean determines the center of the distribution, and the standard deviation determines the spread or girth of the distribution.
- **Common in Nature:** Many natural phenomena follow a Normal distribution. This includes characteristics like people's heights or IQ scores, measurement errors in experiments, and many others.

Given its properties, the Normal distribution serves as a foundation for many statistical procedures and concepts, including hypothesis testing, confidence intervals, and linear regression analysis.

Chapter 3

Change of Variables

Let's say that I'm making ice spheres, and I tell you that the radius of the ice spheres is normally distributed with a mean of 0.7 cm and a standard deviation of 0.08 cm. Then you can draw the probability distribution and cumulative distribution for that:



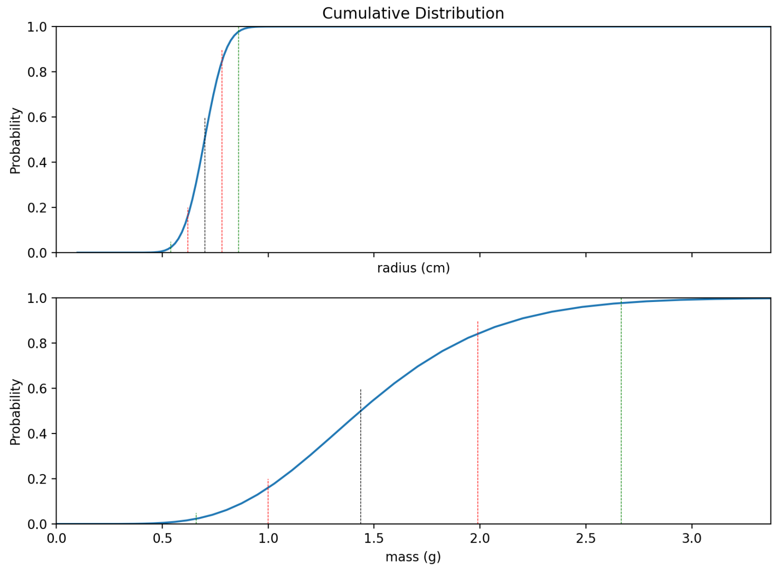
This includes lines indicating the mean and two standard deviations on each side.

Now, let's say I ask you what the cumulative distribution is for the *mass* of the balloons. A cubic centimeter of ice weighs about gram, so if you know the radius of a particular ice sphere, it is easy to compute the mass of it:

$$m = \frac{4}{3}\pi r^3$$

So, for example, if a sphere has a radius of 5cm, its mass in grams is $\frac{4\pi(0.7^3)}{3} \approx 1.44$ g.

Thinking about the graph of the cumulative distribution: if half the balloons have a radius less than 5 cm, than half the balloon have a mass less than 523.6 g. For each point on the cumulative graph, we can use the radius of that point to compute the corresponding mass – the CDF gets stretched out:



If F is the original cumulative distribution function, and g is the function that maps the new variable (mass, in this case) to the old one (radius), then the new cumulative distribution function H is given by

$$H(m) = F(g(m))$$

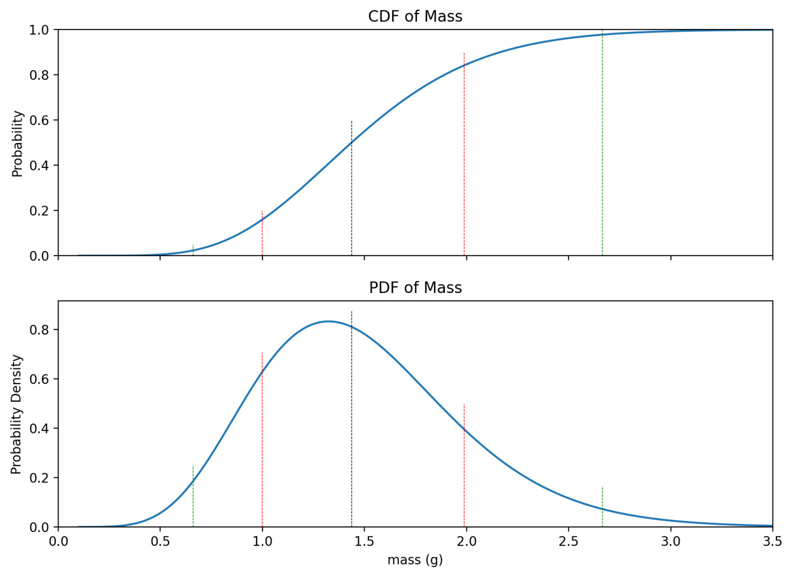
In this case, F is the cumulative function for the normal distribution with mean 0.7 and standard deviation of 0.08. g maps the mass to the radius:

$$g(m) = \left(\sqrt[3]{\frac{3}{4\pi}}\right) x^{\frac{1}{3}}$$

3.1 Making a Probability Density Function

Now we know how to calculate a new cumulative distribution function using the new variable. However, we usually want a probability density.

Here is the CDF and the PDF of the mass of the ice spheres:



Reminder: The probability density function is the derivative of the cumulative distribution function. We know the CDF is

$$H(m) = F(g(m))$$

By the chain rule:

$$H'(m) = F'(g(m))g'(m)$$

The function F is the cumulative distribution for the normal distribution with mean 0.7 and standard deviation of 0.08. So we know its derivative:

$$F'(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Where $\mu = 0.7$ and $\sigma = 0.08$.

We've already said that

$$g(m) = \left(\sqrt[3]{\frac{3}{4\pi}}\right) m^{\frac{1}{3}}$$

Which is easy to differentiate:

$$g'(m) = \left(\frac{1}{3}\right)\left(\sqrt[3]{\frac{3}{4\pi}}\right) m^{-\frac{2}{3}}$$

Here, then, is the code to generate that last plot:

```
import numpy as np
from scipy.stats import norm
import matplotlib.pyplot as plt

# Constants
MEAN_RADIUS = 0.7
STD_RADIUS = 0.08

# Range to plot
MIN_MASS = 0.1
MAX_MASS = 3.5

# Number of points to plot
N = 200

# Needed for radius_for_mass and d_radius_for_mass
C = np.power(3 / (4 * np.pi), 1/3)

# In these three functions, x can
# be a number or a numpy array

def mass_for_radius(x):
    return 4 * np.pi * np.power(x, 3) / 3

def radius_for_mass(x):
    return C * np.power(x, 1/3)

# Derivative of radius_for_mass()
def d_radius_for_mass(x):
    return (C/3) * np.power(x,-2/3)

# Compute mean and 2 standard deviations in each direction
m_mean = mass_for_radius(MEAN_RADIUS)
m_minus_std = mass_for_radius(MEAN_RADIUS - STD_RADIUS)
m_plus_std = mass_for_radius(MEAN_RADIUS + STD_RADIUS)
m_minus_2std = mass_for_radius(MEAN_RADIUS - 2 * STD_RADIUS)
m_plus_2std = mass_for_radius(MEAN_RADIUS + 2 * STD_RADIUS)

# Make N possible values for mass
m_values = np.linspace(MIN_MASS, MAX_MASS, N)

# Compute g(m) for each of these masses
```

```
# That is: What is the radius for each of these masses?
r_values = radius_for_mass(m_values)

# Compute F(g(m)) for each of these masses
# That is: What is the cumulative distribution for each those radii?
cdf_values = norm.cdf(r_values, loc=MEAN_RADIUS, scale=STD_RADIUS)

# Compute g'(m) for each of these masses
dg_values = d_radius_for_mass(m_values)

# What is F'(g(m))g'(m)?
pdf_values = norm.pdf(r_values, loc=MEAN_RADIUS, scale=STD_RADIUS) * dg_values

# Sanity check: It should integrate to a little less then 1.0
dx = (MAX_MASS - MIN_MASS)/N
area_under_curve = pdf_values.sum() * dx
print(f"Integral from {MIN_MASS:.2f} to {MAX_MASS:.2f}: {area_under_curve:.3f}")

# Make a figure with two axes
fig, axs = plt.subplots(nrows=2, sharex=True, figsize=(10, 7), dpi=200)

# Draw the CDF on the second axix
axs[0].set_title("CDF of Mass")
axs[0].set_ylim(bottom=0.0, top=1.0)
axs[0].set_xlim(left=0.0, right=MAX_MASS)
axs[0].set_ylabel("Probability")
axs[0].plot(m_values, cdf_values)

# Add lines for mean,  mean-std, and mean+std
axs[0].vlines(m_minus_2std, 0, 0.05, "g", linestyle="dashed",lw=0.5)
axs[0].vlines(m_minus_std, 0, 0.2, "r", linestyle="dashed",lw=0.5)
axs[0].vlines(m_mean, 0, 0.6, "k", linestyle="dashed",lw=0.5)
axs[0].vlines(m_plus_std, 0, 0.9, "r", linestyle="dashed",lw=0.5)
axs[0].vlines(m_plus_2std, 0, 1.0, "g", linestyle="dashed",lw=0.5)

# How high does the pdf go?
max_density = pdf_values.max()

# Draw the PDF on the second axix
axs[1].set_title("PDF of Mass")
axs[1].set_ylim(bottom=0.0, top=max_density * 1.1)
axs[1].set_xlim(left=0.0, right=MAX_MASS)
axs[1].set_xlabel("mass (g)")
axs[1].set_ylabel("Probability Density")
axs[1].plot(m_values, pdf_values)

# Add lines for mean,  mean-std, and mean+std
axs[1].vlines(m_minus_2std, 0, max_density * .3, "g", linestyle="dashed",lw=0.5)
axs[1].vlines(m_minus_std, 0, max_density * .85 , "r", linestyle="dashed",lw=0.5)
axs[1].vlines(m_mean, 0, max_density * 1.05, "k", linestyle="dashed",lw=0.5)
axs[1].vlines(m_plus_std, 0, max_density * .6, "r", linestyle="dashed",lw=0.5)
axs[1].vlines(m_plus_2std, 0, max_density * .2, "g", linestyle="dashed",lw=0.5)
```

```
fig.savefig("pdf.png")
```

3.2 Decreasing Conversions

The last case (mass and radius) is pretty straightforward because the function g is always increasing. What if we have a change of variables where g is decreasing. For example, $V = IR$ so $\frac{V}{R} = I$.

Let's say that you work at a lightbulb factory and you sample the lightbulbs to see what their resistance is. You find the resistances of the lightbulbs are normally distributed with a mean of 24 ohms and a standard deviation of 3 ohms. The voltage will be exactly 12 volts. What is the PDF of the currents that will pass through the lightbulbs?

$$I = \frac{12}{R}$$

so

$$g(x) = \frac{12}{x}$$

is the function that will convert the current to resistance. Taking the derivative, we get:

$$g'(i) = -\frac{12}{x^2}$$

Chapter 4

Poisson and Exponential Probability Distributions

In this chapter, we will explore two essential probability distributions: the Poisson distribution and the exponential distribution. These distributions play a crucial role in modeling random events and phenomena, providing insights into the occurrence of events over time or in a discrete set of outcomes.

The Poisson distribution is widely used to describe the number of events that occur within a fixed interval of time or space. It is particularly useful when dealing with rare events or events that occur independently at a constant average rate. For example, the Poisson distribution can model the number of customer arrivals at a store in a given hour, the number of phone calls received by a call center in a day, or the number of defects in a production process.

The Poisson distribution is characterized by a single parameter, often denoted as λ, which represents the average rate of event occurrences in the specified interval. The probability mass function of the Poisson distribution gives the probability of observing a specific number of events within that interval.

On the other hand, the exponential distribution is used to model the time between events occurring at a constant average rate. It is commonly employed in reliability analysis, queuing theory, and survival analysis. For example, the exponential distribution can represent the time between customer arrivals at a service desk, the lifespan of electronic components, or the duration between consecutive earthquakes.

The exponential distribution is characterized by a parameter often denoted as λ, which represents the average rate of event occurrence. The probability density function of the exponential distribution describes the likelihood of observing a specific time interval between events.

In this chapter, we will explore the following key aspects of the Poisson and exponential probability distributions:

- Probability mass function and probability density function: We will dive into the mathematical representation of these distributions and learn how to calculate probabilities and densities for specific events or time intervals.
- Mean and variance: We will discuss how to calculate the mean and variance of the

Poisson and exponential distributions, providing measures of central tendency and variability.

- Applications and examples: We will examine real-world scenarios where these distributions find practical applications. From analyzing customer arrival patterns to modeling equipment failure rates, we will explore a range of contexts where the Poisson and exponential distributions prove valuable.
- Relationship between the Poisson and exponential distributions: We will explore the connection between these distributions, as the exponential distribution can emerge as the waiting time between events following a Poisson process.
- Limitations and assumptions: We will also discuss the assumptions and limitations associated with the Poisson and exponential distributions, helping you understand when these models are suitable and when alternative approaches may be necessary.

By developing a solid understanding of the Poisson and exponential probability distributions, you will gain powerful tools for modeling and analyzing random events in various fields. These distributions provide valuable insights into event occurrences, time intervals, and rates, supporting decision-making processes and improving our understanding of stochastic phenomena.

APPENDIX A

Answers to Exercises

Index