Section 1: Probability Review

Molly Roberts

January 26, 2012

Outline

Three Axioms of Probability

Let S be the sample space and A be an event in S.

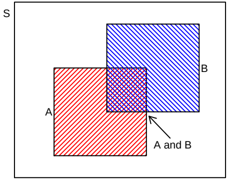
1. For any event A, $P(A) \geq 0$.
   e.g. The probability of getting a heads cannot be -1.
2. $P(S) = 1$.
   e.g. $P(heads) + P(tails) = 1$
3. If $A_1, A_2, \ldots, A_n$ are mutually disjoint, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

e.g. P(rolling a 3 or rolling a 2) $= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

Conditional Probability

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



Conditional Probability

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

Multiplicative Law of Probability:

$$\begin{aligned}
P(A \cap B) &= P(A|B)P(B) \\
P(B \cap A) &= P(B|A)P(A) \\
P(B \cap A) &= P(A \cap B) \\
P(A|B)P(B) &= P(B|A)P(A)
\end{aligned}$$

Bayes' Rule:

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

Side Note: Bayes Rule

Bayes rule will be extremely important going forward.

Why?

- Often we want to know $P(\theta|\text{data})$.
- But what we *do* know is $P(\text{data}|\theta)$.
- We'll be able to *infer* θ by using a variant of Bayes rule. Stay tuned.

Bayes' Rule:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

Outline

Random Variables

A **random variable** is a function from S, the sample space, to R the real line, e.g. When rolling a two dice, we may be interested in whether or not

the sum of the two dice is 7.

Ex. Waiting for the Redline – How long will it take for the next T to get here?

$$X = \begin{cases} 1 & \text{if the redline arrives within 1 minute} \\ 2 & \text{if } 1-2 \text{ minutes} \\ 3 & \text{if } 2-3 \text{ minutes} \\ 4 & \text{if } 3-4 \text{ minutes} \\ \vdots & \vdots \end{cases}$$

Ex. Waiting for the Redline Cont

Now, suppose the probability that the T comes in any given minute is a constant $\pi = .2$, and whether the T comes is independent of what has happened in previous periods.

What's Pr(X=1)?

$$Pr(X = 1) = \pi = .2.$$

What's Pr(X=2)?

$$Pr(X = 2) = (1 - \pi)\pi = .8 \cdot .2 = .16.$$

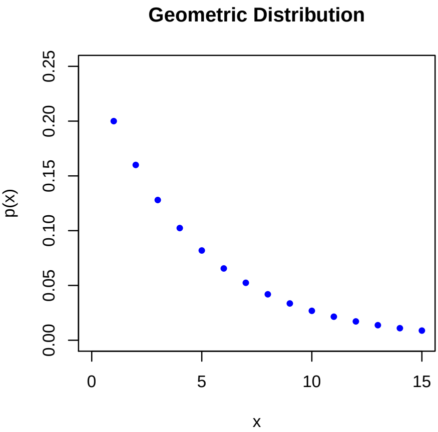
What's Pr(X=3)?

$$Pr(X = 3) = (1 - \pi)^2\pi = .8^2 \cdot .2 = .128$$

And generally...

$$Pr(X = x) = (1 - \pi)^{x-1}\pi = .8^{x-1} \cdot .2$$

A Pictorial Representation of the PMF

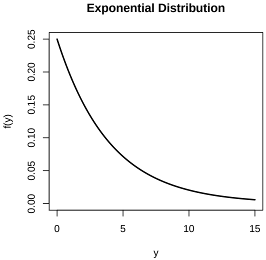


X is a Geometric random variable with parameter $\pi = .2$.

Probability Density Function

Ex. Waiting for the Redline: an alternative model where Y is exact time the T arrives.

$$f(y) = \lambda e^{-\lambda y} = .25e^{-.25y}$$



Y is an Exponential random variable with parameter $\lambda = .25$.

Ex. Calculating probabilities for continuous RVs

Recall that to identify probabilities for continuous random variables we have to use:

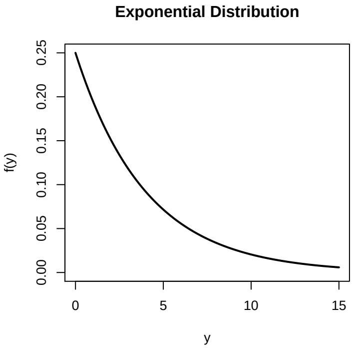
$$P(Y \in A) = \int_A f(y)dy.$$

$$Pr(2 \leq y \leq 10) \quad =$$

Probability Density Function

Ex. Waiting for the Redline.: an alternative model where Y is exact time the T arrives.

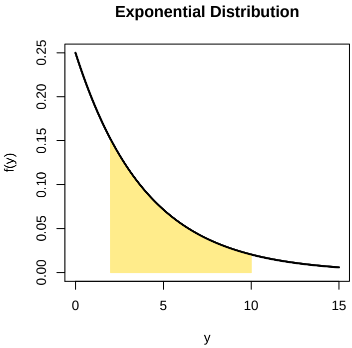
$$f(y) = \lambda e^{-\lambda y} = .25e^{-.25y}$$



Probability Density Function

Ex. Waiting for the Redline.: an alternative model where Y is exact time the T arrives.

$$f(y) = \lambda e^{-\lambda y} = .25e^{-.25y}$$



Ex. Calculating probabilities for continuous RVs

Recall that to identify probabilities we must use[1]:

$$P(Y \in A) = \int_A f(y)dy.$$

$$\begin{aligned} Pr(2 \leq y \leq 10) &= \int_2^{10} .25e^{-.25y} dy \\ &= -e^{-.25y}\Big|_2^{10} \\ &= -e^{-.25 \cdot 10} + e^{-.25 \cdot 2} \\ &\approx .525 \end{aligned}$$

[1] Here we also use the fact that the antiderivative of e^{bx} is $\frac{e^{bx}}{b}$

Characteristics of all PDFs and PMFs:

1. The support is all y's where $P(Y = y) > 0$.
2. Probability mass/density must integrate to 1
3. $P(Y = y) \geq 0$ for all Y

Ex.:

1. $\int_0^\infty .25e^{-.25y}dy = -e^{-.25y}|_0^\infty = 0 + 1 = 1$
2. $.25e^{-.25y} \geq 0$ for all $y \in (0, \infty)$

Expectation

Discrete Case:

$$E(X) = \sum_i x_i P(X = x_i)$$

where $P(X = x)$ is the probability mass function (PMF).

Continuous Case:

$$E(Y) = \int_{-\infty}^{\infty} y f(y) dy$$

where $f(y)$ is the probability density function (PDF).

Our running examples...

Discrete Time:

$$
\begin{aligned}
E(X) &= \sum_i x_i P(X = x_i) \\
&= \sum_{x_i=1}^{\infty} x_i (1 - .2)^{x_i - 1} \cdot .2 \\
&= 5
\end{aligned}
$$

Continuous Time:

$$
\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty} y f(y) dy \\
&= \int_{0}^{\infty} y \cdot .25 e^{-.25y} dy \\
&= 4
\end{aligned}
$$

Expectation of a Function of a Random Variable

Now let's complicate things: suppose we want to find $E[g(X)]$, where $g(X)$ is any function of X.

$$E[g(X)] = \sum_i g(x_i)P(X = x_i)$$

for discrete random variables

$$E[g(y)] = \int_{-\infty}^{\infty} g(y)f(y)dy$$

for continuous random variables.

Return of our running examples

Suppose we want to find $E[g(X)]$, where $g(X) = \sqrt{1+x}$.

$$\begin{aligned} E[g(X)] &= \sum_i g(x_i) P(X = x_i) \\ &= \sum_{x=1}^{\infty} \sqrt{1+x}(1-.2)^{x-1} \cdot .2 \end{aligned}$$

$$\begin{aligned} E[g(Y)] &= \int_{-\infty}^{\infty} g(y) f(y) dx \\ &= \int_{0}^{\infty} \sqrt{1+y}.25e^{-.25y} dy \end{aligned}$$

Outline

The Monte Carlo Method

Basic idea: rather than calculate quantities analytically using deterministic formulae, approximate quantities using random sampling.

Simulating an expectation

$$E(X) = \sum_{x_i=1}^{\infty} x_i(1-.2)^{x_i-1} \cdot .2$$

```
> draws <- rgeom(n = 100000, prob = .2)
```

```
> mean(draws)
[1] 4.99796
```

$$E(Y) = \int_0^{\infty} y \cdot .25e^{-.25y} dy$$

```
> draws <- rexp(n = 100000, rate = .25)
```

```
> mean(draws)
[1] 4.008509
```

Neither of these are perfectly accurate but they become arbitrarily accurate as $n \to \infty$.

Monte Carlo Integration

What we just did was called **Monte Carlo Integration**, which means exactly what it sounds like (doing integrals via Monte Carlo simulation).

If we need to take an integral of the following form:

$$I = \int g(x) f(x) dx$$

Monte Carlo Integration allows us to approximate it by simulating M values from $f(x)$ and calculating:

$$\hat{I}_M = \frac{1}{M} \sum_{i=1}^{M} g(x^{(i)})$$

By the Strong Law of Large Numbers, our estimate $\hat{I}_M$ is a simulation consistent estimator of I as $M \to \infty$(our estimate gets better as we increase the number of simulations).

Back to our examples

$$E[g(X)] = \sum_{x=1}^{\infty} \sqrt{1+x}(1-.2)^{x-1}.2$$

Approach:

1. Draw $M = 100000$ samples from the geometric distribution.
2. Calculate $g(x^{(i)})$ for each.
3. Find the mean of these.

```
> draws <- rgeom(n = 100000, prob = .2)

> g.draws <- sqrt(1 + draws)
> mean(g.draws)
[1] 2.312
```

Back to our examples

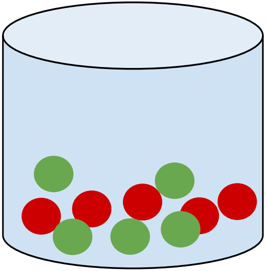
$$E[g(X)] = \int_0^\infty \sqrt{1+y}.25e^{-.25y}dy$$

Approach:

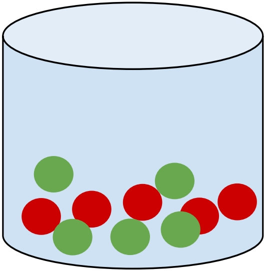
1. Draw $M = 100000$ samples from the exponential distribution.
2. Calculate $g(x^{(i)})$ for each.
3. Find the mean of these.

```
> draws <- rexp(n = 100000, rate = .25)
> g.draws <- sqrt(1 + draws)
> mean(g.draws)
[1] 2.096
```

Simulation for Probability Problems



I have an urn composed of 5 red balls and 5 green balls. If I sample 4 balls without replacement from the urn, what is the probability of drawing 4 balls all of the same color?



Here are some steps:

1. Construct our population aka our urn.
2. Figure out how to take one sample from it.
3. Figure out a programming rule for determining whether our condition was met i.e. we drew "4 red balls or 4 green balls".
4. Throw a for loop around it and sample repeatedly.
5. Determine what proportion of times our condition was successful.

1. Construct our population

Here are some possibilities:

```
> urn <- c("G","G","G","G","G","R","R","R","R","R")
> urn
 [1] "G" "G" "G" "G" "G" "R" "R" "R" "R" "R"
```

```
> urn <- c(rep("red",5),rep("green",5))
> urn
 [1] "red"   "red"   "red"   "red"   "red"
  "green" "green" "green" "green"  "green"
```

```
> urn <- c(rep(1,5),rep(0,5))
> urn
 [1] 1 1 1 1 1 0 0 0 0 0
```

I'll use this last one because numbers will be easier to use later on.

2. Figure out how to take one sample

We need to use the `sample()` function with the `replace = FALSE` argument:

```
> sample(x = urn, size = 4, replace = FALSE)
[1] 1 1 1 0
```

3. Determine if a success or failure

Because we used numeric signifiers for red and green, there is an easy test for whether or not we have drawn balls of all one color. If the numbers sum up to either 0 or 4, then we have a 'success'.

```
> draw <- sample(x = urn, size = 4, replace = FALSE)
> draw
[1] 0 0 0 0

> sum(draw) == 4
[1] FALSE
> sum(draw) == 0
[1] TRUE
```

But we can combined these test using '|' which means 'or'

```
> sum(draw == 4) | sum(draw == 0)
[1] TRUE
```

4. Throw a for-loop around it

Here's our guts so far:

```
draw <- sample(x = urn, size = 4, replace = FALSE)
success <- sum(draw == 4) | sum(draw == 0)
```

We can repeat this ad nauseum:

```
sims <- 1000
success <- NULL
for(i in 1:sims){
  draw <- sample(x = urn, size = 4, replace = FALSE)
  success[i] <- sum(draw) == 4 | sum(draw) == 0
}
```

5. Determine proportion of success

Here are two equivalent approaches.

```
> sum(success)/sims
[1] 0.047
> mean(success)
[1] 0.047
```

Why is this useful?

- Math is hard or impossible and takes too long.
- Consider trying to integrate

$$I(f) = \frac{1}{2\pi} \int_0^1 e^{-\frac{x^2}{2}} dx$$

  Which is the standard normal density and cannot be evaluated in closed form.
- How could we solve this?
  1. Sample 1000 points, $X_1, \ldots, X_{1000}$, uniformly distributed over the interval $(0, 1)$.
  2. Evaluate the function at each of these points and take the mean.

$$\frac{1}{1000}\left(\frac{1}{2\pi}\right)\sum_{i=1}^{1000} e^{-\frac{x_i^2}{2}}$$