Probability and Random Variables

Concepts

1. Utility and Probability
2. Random Variables
3. Relationships between Random Variables

Utility and Probability

Full story: https://projecteuclid.org/journals/statistical-science/volume-1/issue-3/The-Axioms-of-Subjective-Probability/10.1214/ss/1177013611.full

Utility and Probability

Consider events A and B:

Full story: https://projecteuclid.org/journals/statistical-science/volume-1/issue-3/The-Axioms-of-Subjective-Probability/10.1214/ss/1177013611.full

Utility and Probability

Consider events A and B:

<u>Utility indicates preference</u>

Full story: https://projecteuclid.org/journals/statistical-science/volume-1/issue-3/The-Axioms-of-Subjective-Probability/10.1214/ss/1177013611.full

Utility and Probability

Consider events A and B:

Utility indicates preference

Probability indicates plausibility

Full story: https://projecteuclid.org/journals/statistical-science/volume-1/issue-3/The-Axioms-of-Subjective-Probability/10.1214/ss/1177013611.full

Utility and Probability

Consider events A and B:

Utility indicates preference

$U(A) > U(B)$	Indicates A is *preferable* to B
$U(A) = U(B)$	Indicates *indifference* between A and B

Probability indicates plausibility

Full story: https://projecteuclid.org/journals/statistical-science/volume-1/issue-3/The-Axioms-of-Subjective-Probability/10.1214/ss/1177013611.full

Utility and Probability

Consider events A and B:

Utility indicates preference

$U(A) > U(B)$	Indicates A is *preferable* to B
$U(A) = U(B)$	Indicates *indifference* between A and B

Probability indicates plausibility

$P(A) > P(B)$	Indicates A is *more plausible* (or likely) than B
$P(A) = P(B)$	Indicates A is *equally as plausible* (or equally likely) as B

Full story: https://projecteuclid.org/journals/statistical-science/volume-1/issue-3/The-Axioms-of-Subjective-Probability/10.1214/ss/1177013611.full

What is a Random Variable?

R.V. X

Vocabulary/Notation

Vocabulary/Notation

Term	Definition	Coinflip Example

Vocabulary/Notation

Term	Definition	Coinflip Example
support(X)		

Vocabulary/Notation

Term	Definition	Coinflip Example
support(X)	All the values that X can take	

Vocabulary/Notation

Term	Definition	Coinflip Example
support(X) $x \in X$	All the values that X can take	

Vocabulary/Notation

Term	Definition	Coinflip Example
support(X) $x \in X$	All the values that X can take	$\{h, t\}$ or $\{0, 1\}$ "Binary random variable"

Vocabulary/Notation

Term	Definition	Coinflip Example
support(X) $x \in X$	All the values that X can take	$\{h, t\}$ or $\{0, 1\}$ "Binary random variable"
Distribution		

Vocabulary/Notation

Term	Definition	Coinflip Example
support(X) $x \in X$	All the values that X can take	$\{h, t\}$ or $\{0, 1\}$ "Binary random variable"
Distribution	Maps each value in the support to a real number indicating its probability	

Vocabulary/Notation

Term	Definition	Coinflip Example
support(X) $x \in X$	All the values that X can take	$\{h, t\}$ or $\{0, 1\}$ "Binary random variable"
Distribution	Maps each value in the support to a real number indicating its probability	$\mathrm{Bernoulli}(0.6)$ $P(X = 1) = 0.6$ $P(X = 0) = 0.4$

Vocabulary/Notation

Term	Definition	Coinflip Example
support(X) $x \in X$	All the values that X can take	$\{h, t\}$ or $\{0, 1\}$ "Binary random variable"
Distribution	Maps each value in the support to a real number indicating its probability	$\mathrm{Bernoulli}(0.6)$ $P(X = 1) = 0.6$ $P(X = 0) = 0.4$

$P(X)$ is a table

X	P(X)
0	0.4
1	0.6

Vocabulary/Notation

Term	Definition	Coinflip Example	
support(X) $x \in X$	All the values that X can take	$\{h, t\}$ or $\{0, 1\}$ "Binary random variable"	
Distribution	Maps each value in the support to a real number indicating its probability	$\text{Bernoulli}(0.6)$ $P(X = 1) = 0.6$ $P(X = 0) = 0.4$	$P(X)$ is a table
Expectation			

X	P(X)
0	0.4
1	0.6

Vocabulary/Notation

Term	Definition	Coinflip Example	
support(X) $x \in X$	All the values that X can take	$\{h, t\}$ or $\{0, 1\}$ "Binary random variable"	
Distribution	Maps each value in the support to a real number indicating its probability	$\mathrm{Bernoulli}(0.6)$ $P(X = 1) = 0.6$ $P(X = 0) = 0.4$	$P(X)$ is a table
Expectation	First moment of the random variable, "mean"		

X	P(X)
0	0.4
1	0.6

Vocabulary/Notation

Term	Definition	Coinflip Example	
support(X) $x \in X$	All the values that X can take	$\{h, t\}$ or $\{0, 1\}$ "Binary random variable"	
Distribution	Maps each value in the support to a real number indicating its probability	$\text{Bernoulli}(0.6)$ $P(X = 1) = 0.6$ $P(X = 0) = 0.4$	$P(X)$ is a table
Expectation $E[X]$	First moment of the random variable, "mean"		

X	P(X)
0	0.4
1	0.6

Vocabulary/Notation

Term	**Definition**	**Coinflip Example**	
support(X) $x \in X$	All the values that X can take	$\{h, t\}$ or $\{0, 1\}$ "Binary random variable"	
Distribution	Maps each value in the support to a real number indicating its probability	$\text{Bernoulli}(0.6)$ $P(X = 1) = 0.6$ $P(X = 0) = 0.4$	$P(X)$ is a table
Expectation $E[X]$	First moment of the random variable, "mean"	$E[X] = \sum_{x \in X} xP(x)$	

X	**P(X)**
0	0.4
1	0.6

Vocabulary/Notation

Term	**Definition**	**Coinflip Example**	
support(X) $x \in X$	All the values that X can take	$\{h, t\}$ or $\{0, 1\}$ "Binary random variable"	
Distribution	Maps each value in the support to a real number indicating its probability	$\mathrm{Bernoulli}(0.6)$ $P(X = 1) = 0.6$ $P(X = 0) = 0.4$	$P(X)$ is a table
Expectation $E[X]$	First moment of the random variable, "mean"	$E[X] = \sum_{x \in X} xP(x)$ $= 0.5$	

X	**P(X)**
0	0.4
1	0.6

Distributions of related R.V.s

Distributions of related R.V.s

Joint Distribution

Distributions of related R.V.s

Joint Distribution

$$P(X, Y, Z)$$

Distributions of related R.V.s

Joint Distribution

$$P(X, Y, Z)$$

X	Y	Z	$P(X,Y,Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

Distributions of related R.V.s

Joint Distribution

$$P(X, Y, Z)$$

X	Y	Z	$P(X,Y,Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

Conditional Distribution

Distributions of related R.V.s

Joint Distribution

$$P(X, Y, Z)$$

X	Y	Z	$P(X, Y, Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

Conditional Distribution

$$P(X \mid Y, Z)$$

Distributions of related R.V.s

Joint Distribution

$$P(X, Y, Z)$$

X	Y	Z	$P(X, Y, Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

Conditional Distribution

$$P(X \mid Y, Z)$$

(Distribution - valued function)

X	$P(X \mid Y=1, Z=1)$
0	0.84
1	0.16

Distributions of related R.V.s

Joint Distribution

$$P(X, Y, Z)$$

X	Y	Z	$P(X, Y, Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

Conditional Distribution

$$P(X \mid Y, Z)$$

(Distribution - valued function)

X	$P(X \mid Y=1, Z=1)$
0	0.84
1	0.16

Marginal Distribution

Distributions of related R.V.s

Joint Distribution

$$P(X, Y, Z)$$

X	Y	Z	$P(X, Y, Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

Conditional Distribution

$$P(X \mid Y, Z)$$

(Distribution - valued function)

X	$P(X \mid Y=1, Z=1)$
0	0.84
1	0.16

Marginal Distribution

$$P(X)\ P(Y)\ P(Z)$$

Distributions of related R.V.s

Joint Distribution

$$P(X, Y, Z)$$

X	Y	Z	$P(X,Y,Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

Conditional Distribution

$$P(X \mid Y, Z)$$

(Distribution - valued function)

X	$P(X \mid Y=1, Z=1)$
0	0.84
1	0.16

Marginal Distribution

$$P(X)\ P(Y)\ P(Z)$$

X	$P(X)$
0	0.85
1	0.15

Y	$P(Y)$
0	0.45
1	0.55

Z	$P(Z)$
0	0.20
1	0.80

Distributions of related R.V.s

Joint Distribution

$P(X, Y, Z)$

Conditional Distribution

$P(X \mid Y, Z)$

Marginal Distribution

$P(X)\ P(Y)\ P(Z)$

Distributions of related R.V.s

Joint Distribution

$P(X, Y, Z)$

Conditional Distribution

$P(X \mid Y, Z)$

Marginal Distribution

$P(X)\ P(Y)\ P(Z)$

3 Rules

Distributions of related R.V.s

Joint Distribution

$P(X, Y, Z)$

Conditional Distribution

$P(X \mid Y, Z)$

Marginal Distribution

$P(X)\ P(Y)\ P(Z)$

3 Rules

(Burrito-level)

Distributions of related R.V.s

Joint Distribution

$$P(X, Y, Z)$$

Conditional Distribution

$$P(X \mid Y, Z)$$

Marginal Distribution

$$P(X)\ P(Y)\ P(Z)$$

3 Rules

(Burrito-level)

(Gourmet Level: Axioms of Probability)

not on exam

AXIOM 1. STRUCTURE OF UNKNOWN REAL NUMBERS AND PLAUSIBLE VALUE. *We assume a set T of unknown numbers is a partially ordered commutative algebra over $\mathbb{R}$ with identity, 1.*

We assume in addition a given sub-Boolean algebra E of $E(T)$ with $0, 1 \in E$ and denote by E_0 the set of non-zero members of E. We assume that the partial ordering in $E(T)$ as a Boolean algebra coincides with the ordering that $E(T)$ inherits from the algebra T. Finally, we assume a function $PV : T \times E_0 \to \mathbb{R}$, called **PLAUSIBLE VALUE**, *whose value on the pair (x, e) is denoted $PV(x|e)$.*

AXIOM 2. STRONG RESCALING FOR PLAUSIBLE VALUE. *If a, b belong to $\mathbb{R}$, if x belongs to T, and if e belongs to E_0, then*

$$PV(ax + b|e) = aPV(x|e) + b. \tag{2}$$

AXIOM 3. ORDER CONSISTENCY FOR PLAUSIBLE VALUE. *If $x, y \in T$ and if $e \in E_0$, implies that $x \le y$, then $PV(x|e) \le PV(y|e)$.*

Notice that if $e \in E(T)$, then $0 \le e \le 1$, in T, as it is true in the lattice ordering of $E(T)$.

AXIOM 4. THE COX AXIOM FOR PLAUSIBLE VALUE: *If e, c are fixed in E, with $ec \in E_0$, if x_1, x_2 are in T, if $PV(x_1|ec) = PV(x_2|ec)$, then $PV(x_1e|c) = PV(x_2e|c)$. That is, we assume that as a function of x, the plausible value $PV(xe|c)$ depends only on $PV(x|ec)$.*

AXIOM 5. RESTRICTED ADDITIVITY OF PLAUSIBLE VALUE. *For each fixed $y \in T$ and $e \in E_0$, the plausible value $PV(x + y|e)$ as a function of $x \in T$ depends only on $PV(x|e)$, which is to say that if $x_1, x_2 \in T$ and $PV(x_1|e) = PV(x_2|e)$, then $PV(x_1 + y|e) = PV(x_2 + y|e)$.*

Distributions of related R.V.s

Joint Distribution

$P(X, Y, Z)$

Conditional Distribution

$P(X \mid Y, Z)$

Marginal Distribution

$P(X)\ P(Y)\ P(Z)$

3 Rules

(Burrito-level)

1)

Distributions of related R.V.s

Joint Distribution

$P(X, Y, Z)$

Conditional Distribution

$P(X \mid Y, Z)$

Marginal Distribution

$P(X)\ P(Y)\ P(Z)$

3 Rules

(Burrito-level)

1) a) $0 \leq P(X \mid Y) \leq 1$

Distributions of related R.V.s

Joint Distribution

$P(X, Y, Z)$

Conditional Distribution

$P(X \mid Y, Z)$

Marginal Distribution

$P(X)\ P(Y)\ P(Z)$

3 Rules

(Burrito-level)

1) a) $0 \leq P(X \mid Y) \leq 1$

   b) $\sum_{x \in X} P(x \mid Y) = 1$

Distributions of related R.V.s

Joint Distribution

$P(X, Y, Z)$

Conditional Distribution

$P(X \mid Y, Z)$

Marginal Distribution

$P(X)\ P(Y)\ P(Z)$

3 Rules

(Burrito-level)

1) a) $0 \leq P(X \mid Y) \leq 1$

   b) $\sum_{x \in X} P(x \mid Y) = 1$

2) "Law of total probability"

$$P(X) = \sum_{y \in Y} P(X, y)$$

Distributions of related R.V.s

Joint Distribution

$P(X, Y, Z)$

Conditional Distribution

$P(X \mid Y, Z)$

Marginal Distribution

$P(X)\ P(Y)\ P(Z)$

3 Rules

(Burrito-level)

1) a) $0 \leq P(X \mid Y) \leq 1$

   b) $\sum_{x \in X} P(x \mid Y) = 1$

2) "Law of total probability"

$$P(X) = \sum_{y \in Y} P(X, y)$$

Joint $\rightarrow$ Marginal

Distributions of related R.V.s

Joint Distribution

$P(X, Y, Z)$

Conditional Distribution

$P(X \mid Y, Z)$

Marginal Distribution

$P(X)\ P(Y)\ P(Z)$

3 Rules

(Burrito-level)

1) a) $0 \leq P(X \mid Y) \leq 1$

   b) $\sum_{x \in X} P(x \mid Y) = 1$

2) "Law of total probability"

$$P(X) = \sum_{y \in Y} P(X, y)$$

Joint $\rightarrow$ Marginal

3) Definition of Conditional Probability

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

Distributions of related R.V.s

Joint Distribution

$P(X, Y, Z)$

Conditional Distribution

$P(X \mid Y, Z)$

Marginal Distribution

$P(X)\ P(Y)\ P(Z)$

3 Rules

(Burrito-level)

1) a) $0 \leq P(X \mid Y) \leq 1$

   b) $\sum_{x \in X} P(x \mid Y) = 1$

2) "Law of total probability"

$$P(X) = \sum_{y \in Y} P(X, y)$$

3) Definition of Conditional Probability

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

Joint $\rightarrow$ Marginal

Joint + Marginal $\rightarrow$ Conditional

Distributions of related R.V.s

Joint Distribution | **Conditional Distribution** | **Marginal Distribution**

$P(X, Y, Z)$ | $P(X \mid Y, Z)$ | $P(X)\, P(Y)\, P(Z)$

3 Rules

(Burrito-level)

1) a) $0 \leq P(X \mid Y) \leq 1$
   b) $\sum_{x \in X} P(x \mid Y) = 1$
2) "Law of total probability"

$$P(X) = \sum_{y \in Y} P(X, y)$$

3) Definition of Conditional Probability

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

Joint $\rightarrow$ Marginal

Joint + Marginal $\rightarrow$ Conditional

Marginal + Conditional $\rightarrow$ Joint

$$P(X, Y) = P(X|Y)\, P(Y)$$

Distributions of related R.V.s

Joint Distribution

$P(X, Y, Z)$

Conditional Distribution

$P(X \mid Y, Z)$

Marginal Distribution

$P(X)\ P(Y)\ P(Z)$

X	Y	Z	$P(X,Y,Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

3 Rules

1) a) $0 \leq P(X \mid Y) \leq 1$

   b) $\sum_{x \in X} P(x \mid Y) = 1$

2) "Law of total probability"

$$P(X) = \sum_{y \in Y} P(X, y)$$

3) Definition of Conditional Probability

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

Joint $\rightarrow$ Marginal

Joint + Marginal $\rightarrow$ Conditional

Marginal + Conditional $\rightarrow$ Joint

$$P(X, Y) = P(X|Y)\, P(Y)$$

Naive Inference

(Book introduces unnormalized "factors", but process is the same.)

Naive Inference

Three Random Variables: A, B, C (Works for any number)

(Book introduces unnormalized "factors", but process is the same.)

Naive Inference

Three Random Variables: A, B, C (Works for any number)

Want to find $P(\underbrace{A = a}_{\text{query}} \mid \underbrace{B = b}_{\text{evidence}})$

(Book introduces unnormalized "factors", but process is the same.)

Naive Inference

Three Random Variables: A, B, C (Works for any number)

Want to find $P(\underbrace{A=a}_{\text{query}} \mid \underbrace{B=b}_{\text{evidence}})$

C is a "hidden variable"

(Book introduces unnormalized "factors", but process is the same.)

Naive Inference

Three Random Variables: A, B, C (Works for any number)

Want to find $P(\underbrace{A = a}_{\text{query}} \mid \underbrace{B = b}_{\text{evidence}})$

C is a "hidden variable"

1. Determine the joint distribution $P(A, B, C)$.

(Book introduces unnormalized "factors", but process is the same.)

Naive Inference

Three Random Variables: A, B, C (Works for any number)

Want to find $P(\underbrace{A = a}_{\text{query}} \mid \underbrace{B = b}_{\text{evidence}})$

C is a "hidden variable"

1. Determine the joint distribution $P(A, B, C)$.
2. Marginalize over hidden and query variables to get

$$P(A = a, B = b) = \sum_{c} P(A = a, B = b, C = c)$$

and

$$P(B = b) = \sum_{a,c} P(A = a, B = b, C = c)$$

(Book introduces unnormalized "factors", but process is the same.)

Naive Inference

Three Random Variables: A, B, C (Works for any number)

Want to find $P(\underbrace{A=a}_{\text{query}} \mid \underbrace{B=b}_{\text{evidence}})$

C is a "hidden variable"

1. Determine the joint distribution $P(A, B, C)$.
2. Marginalize over hidden and query variables to get

$$P(A=a, B=b) = \sum_c P(A=a, B=b, C=c)$$

and

$$P(B=b) = \sum_{a,c} P(A=a, B=b, C=c)$$

3. $P(A=a \mid B=b) = \frac{P(A=a, B=b)}{P(B=b)}$

(Book introduces unnormalized "factors", but process is the same.)

1) a) $0 \leq P(X \mid Y) \leq 1$

b) $\sum_{x \in X} P(x \mid Y) = 1$

2) $P(X) = \sum_{y \in Y} P(X, y)$

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

$$P(X, Y) = P(X|Y)\, P(Y)$$

Break

- $P \in \{0, 1\}$: Powder Day
- $C \in \{0, 1\}$: Pass Clear
- 1 in 5 days is a powder day
- The pass is clear 8 in 10 days
- If it is a powder day, there is a 50% chance the pass is blocked

- Write out the joint probability distribution for P and C.
- Suppose it is a non-powder day, what is the probability that the pass is blocked?

Bayes Rule

- Know: $P(B \mid A)$, $P(A)$, $P(B)$
- Want: $P(A \mid B)$

Definitions: Conditional Expectation and Independence

Definitions: Conditional Expectation and Independence

Definition: The conditional expectation of X given Y is

$$E[X \mid Y] = \sum_x x\, P(X = x \mid Y)$$

(function from values of Y to expectations of X)

Definitions: Conditional Expectation and Independence

Definition: The conditional expectation of X given Y is

$$E[X \mid Y] = \sum_x x\, P(X = x \mid Y)$$

(function from values of Y to expectations of X)

Definition: X and Y are *independent* iff $P(X, Y) = P(X)\, P(Y)$

Definitions: Conditional Expectation and Independence

Definition: The conditional expectation of X given Y is

$$E[X \mid Y] = \sum_x x\, P(X = x \mid Y)$$

(function from values of Y to expectations of X)

Definition: X and Y are *independent* iff $P(X, Y) = P(X)\, P(Y)$

$X \perp Y$

Definitions: Conditional Expectation and Independence

Definition: The conditional expectation of X given Y is

$$E[X \mid Y] = \sum_x x\, P(X = x \mid Y)$$

(function from values of Y to expectations of X)

Definition: X and Y are *independent* iff $P(X, Y) = P(X)\, P(Y)$

$X \perp Y$

$P(X \mid Y) = P(X)$

Definitions: Conditional Expectation and Independence

Definition: The conditional expectation of X given Y is

$$E[X \mid Y] = \sum_x x \, P(X = x \mid Y)$$

(function from values of Y to expectations of X)

Definition: X and Y are *independent* iff $P(X, Y) = P(X)\, P(Y)$

$X \perp Y$

$P(X \mid Y) = P(X)$

Definition: X and Y are *conditionally independent* given Z iff
$P(X, Y \mid Z) = P(X \mid Z)\, P(Y \mid Z)$

Definitions: Conditional Expectation and Independence

Definition: The conditional expectation of X given Y is

$$E[X \mid Y] = \sum_x x\, P(X = x \mid Y)$$

(function from values of Y to expectations of X)

Definition: X and Y are *independent* iff $P(X, Y) = P(X)\, P(Y)$

$X \perp Y$

$P(X \mid Y) = P(X)$

Definition: X and Y are *conditionally independent* given Z iff
$P(X, Y \mid Z) = P(X \mid Z)\, P(Y \mid Z)$

$X \perp Y \mid Z$

Concepts

1. Utility and Probability
2. Random Variables
3. Relationships between Random Variables