10-701 Basic Probability Recitation

Dan Schwartz

2017-01-19

Events, Event Spaces

- Events (here denoted ω) are possible outcomes of a random experiment
- An event space (here Ω) is the set of all possible outcomes
- Examples:
  - Coin toss: ω = $Heads$
  - Coin toss: ω = $Tails$
  - Coin toss: $\Omega = \{Heads, Tails\}$
  - Die: $\omega \in \Omega = \{1,2,3,4,5,6\}$
  - Height: $\omega \in [0, \infty)$

Random Variables

- Random variables are just functions from events to the real numbers:

$$X : \Omega \to \mathbb{R}$$

Examples:

$$X(True) = 1, X(False) = 0$$

$$X(1) = 1$$

$$X(h) \coloneqq h^2$$

Probability

Probability measure is in reference to a subset of event outcomes occurring.

It is a function from subsets to $[0, 1]$. It can be helpful to think of this as the area covered by a set of events within the total area of the event space.

The notation can be confusing:

$\mathrm{P}(X = 5)$ means $\mathrm{P}(\{\omega\} : X(\omega) = 5) = \mathrm{P}\big(\omega \in X^{-1}(5)\big)$

$\mathrm{P}(X \leq 5)$ means $\mathrm{P}(\{\omega\} : X(\omega) \leq 5) = \mathrm{P}\big(\omega \in X^{-1}((-\infty, 5])\big)$

Other forms:

$\mathrm{P}(X)$ often means $\mathrm{P}(X = x)$

$\mathrm{P}(A)$ can mean $\mathrm{P}(\omega \in A)$ here A denotes a subset not a random variable

Axioms

$$\mathrm{P}(\Omega) = 1$$

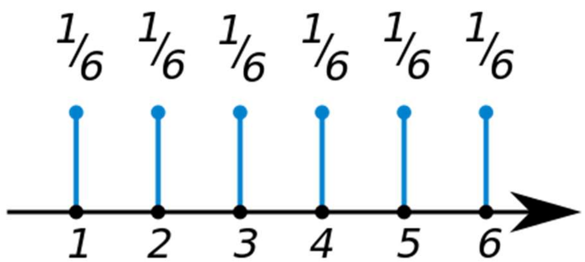
$$\mathrm{P}(\emptyset) = 0$$

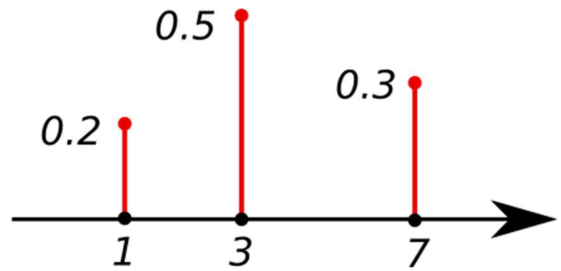
$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B), \text{for A disjoint from B}$$

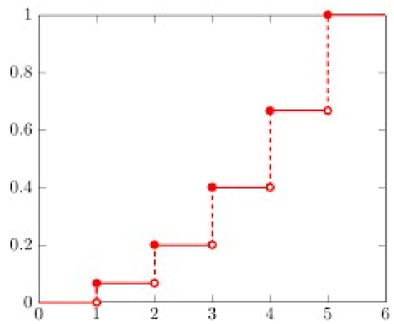
Distributions

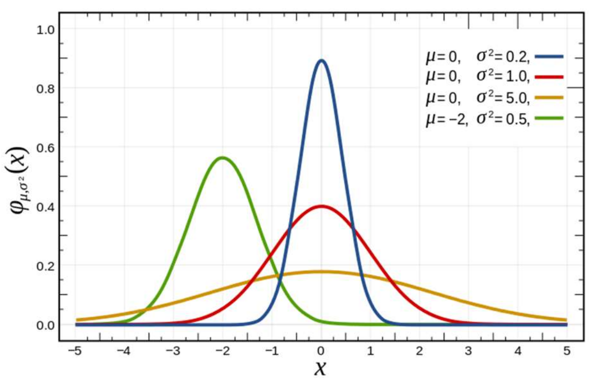
- Discrete
  - probability mass function (pmf) gives $\mathrm{P}(\{\omega_i\})$ for each outcome
- Continuous
  - cumulative distribution function (cdf), denoted $\mathrm{F}(t)$,
    gives $\mathrm{P}(X \leq t)$ for $t \in \mathbb{R}$
    $\mathrm{F}(-\infty) = 0, \mathrm{F}(\infty) = 1$
  - if there exists f, such that $\int_{-\infty}^{t} \mathrm{f}(x)\,\mathrm{d}x = P(X \leq t)$, then f is called the probability density function (pdf)
  - pdf is not the probability of the input value
    - Ratio of the pdf of 2 values can be interpreted as relative likelihood
    - The integral between 2 values is the probability of those values occurring
- Cumulative distribution function applies to both discrete and continuous event spaces

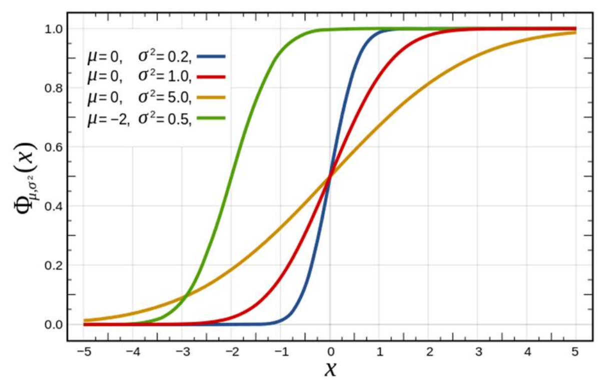
pmf, pdf, cdf examples











median, mode

Median

$t : \mathrm{F}(t) = 0.5, \mathrm{P}(X \leq t) = \mathrm{P}(X > t)$

Mode

point where pmf or pdf is maximum

mean, variance

Mean (also called the expected value)

weighted average of $X(\omega)$, where the weights are given by the probability measure, e.g.:

$$\mathbb{E}[X] = 1 * \mathrm{P}(X = 1) + 0 \ * \mathrm{P}(X = 0) = 0.5$$

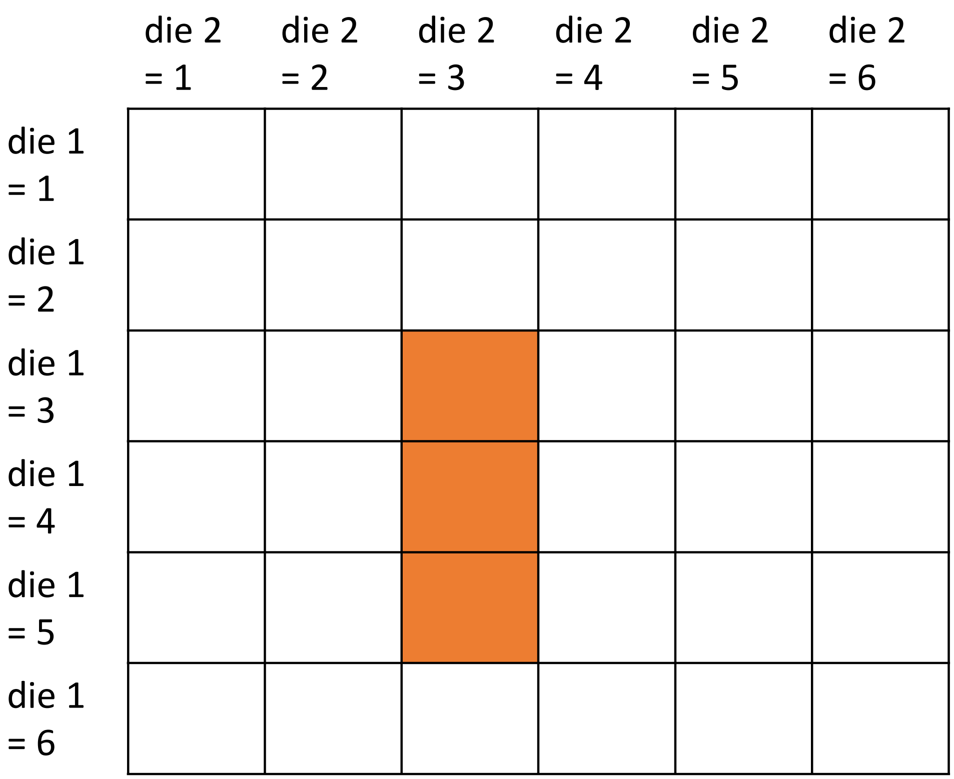
For continuous event spaces: $\mathbb{E}[X] = \int x \ \mathrm{f}(x)\mathrm{d}x$

Can take the expectation of functions: $\mathbb{E}[\mathrm{g}(X)] = \int \mathrm{g}(x) \ \mathrm{f}(x)\mathrm{d}x$

Variance: $\mathbb{E}[(X - \mathbb{E}[X])^2]$, how far do values tend to be from the mean, measure of dispersion

Joint distributions

$$\mathrm{P}_{XY}(X = x, Y = y)$$



Multidimensional event space, we consider an event to be an outcome for all of the variables jointly. $\omega = (x, y)$ is a tuple, or vector.

cdf is also joint:

$$\mathrm{F}_{XY}(x, y) = \mathrm{P}(X \leq x, Y \leq y)$$

$$\mathrm{P}_{XY}(X \in \{3, 4, 5\}, Y \in \{3\})$$

Marginal distributions

$$\mathrm{P}_X(X = x) = \sum_y \mathrm{P}_{XY}(X = x, Y = y)$$

	die 2 = 1	die 2 = 2	die 2 = 3	die 2 = 4	die 2 = 5	die 2 = 6
die 1 = 1						
die 1 = 2						
die 1 = 3						
die 1 = 4						
die 1 = 5						
die 1 = 6						

$\mathrm{P}_X(X = 2)$

Continuous case:

$$\mathrm{f}_X = \int_y \mathrm{f}_{XY}\,(x, y)\mathrm{d}y$$

Independence:

X, Y are independent iff

$$\mathrm{P}_{XY}(X = x, Y = y) = \mathrm{P}_X(X = x)\mathrm{P}_Y(Y = y)\ \forall\, x, y$$

mean, covariance

If $Z = (X, Y)$, then $\mathbb{E}[Z] = (\mathbb{E}[X], \mathbb{E}[Y])$

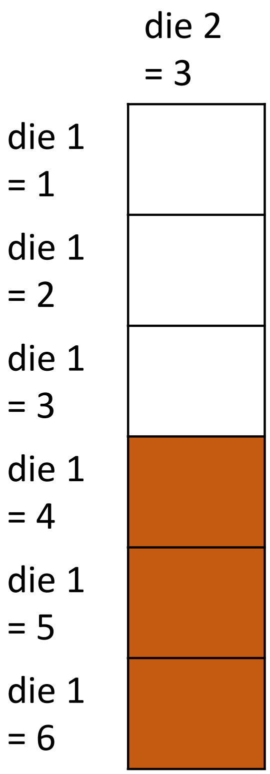
$$\mathbb{E}[\mathrm{g}(Z)] = \int_x \int_y \mathrm{g}((X, Y))\, \mathrm{d}x\, \mathrm{d}y$$

covariance: $\mathbb{E}\left[\left(Z - \mathbb{E}(Z)\right)\left(Z - \mathbb{E}(Z)\right)^{\mathrm{T}}\right]$

How do pairs of variables vary with each other

Conditional distributions, Bayes rule

$\mathrm{P}(X|Y) = \frac{\mathrm{P}(X,Y)}{\mathrm{P}(Y)}$ means the same thing as $\mathrm{P}(X = s|Y = t) = \frac{\mathrm{P}(X=s,Y=t)}{\mathrm{P}(Y=t)}$



$P_{XY}(X \geq 4 \,|Y = 3)$

$$\mathrm{P}(X|Y)\mathrm{P}(Y) = \mathrm{P}(X,Y) = \mathrm{P}(Y|X)\mathrm{P}(X)$$

$$\mathrm{P}(X|Y)\mathrm{P}(Y) = \mathrm{P}(Y|X)\mathrm{P}(X)$$

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

$$\frac{\mathrm{P}(X|Y)\mathrm{P}(Y)}{\sum_y \mathrm{P}(X = x|Y = y)\mathrm{P}(Y = y)} = \mathrm{P}(Y|X)$$

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

Continuous case, conditional independence

$$\mathrm{f}_X(x|Y=y) = \frac{\mathrm{f}_Y(y|X=x)\mathrm{f}_X(x)}{\mathrm{f}_Y(y)}$$

Two variables X, Y are conditionally independent given Z iff:

$$\mathrm{P}(X,Y|Z) = \mathrm{P}(X|Z)\mathrm{P}(Y|Z)\ \forall\ x, y$$

Prior, Likelihood, Posterior

Typical situation: We have a vector of parameters, Θ, in our model and a vector of observations D. We would like to estimate the "best" parameters for our model, by maximizing the probability of the conditional distribution $\mathrm{P}(\Theta|D)$. In this case we name the parts of our conditional distribution:

$$\mathrm{P}(\Theta|D) = \frac{\mathrm{P}(D|\Theta)\mathrm{P}(\Theta)}{\mathrm{P}(D)}$$

Example: We assume the data follow a Gaussian distribution, i.e. the likelihood is a Gaussian. We try to estimate Θ, the mean and variance of the Gaussian, using the posterior

Distribution Families

- We often assume various families to keep the math tractable:
  - Gaussian for regression, Gaussian mixtures, etc.
  - Categorical/Multinomial for discrete data