Probability Distributions
Probability Distributions

A random variable is a symbolic name given to an experimental outcome, e.g. $X$ is the face of a coin following a toss.

A probability distribution is an a description of probability measures to values a random variable may take.

A distribution is *discrete* if the set of outcomes is *enumerable*

- e.g. can be put in 1-1 correspondence with non-negative integers

- Examples
  
  - $\Omega = \{1,2,3,4,5,6\}$ (faces of a die)
  
  - $\Omega = \{TTT, TTH, THT, THH, HTT,HTH,HHT,HHH\}$ (3 coin flips)
  
  - $\Omega = \{2,4,6,8,10, \ldots \}$ (all even integers)
  
  - $\Omega = \{(0,0,0,0), (0,0,0,1), (0,0,0,2), \ldots \}$ all four-tuples of non-negative integers
Discrete Probability Distribution (1)

To be in 1-1 correspondence with non-negative integers means there is a 1-1 function

\[ f: \Omega \rightarrow \{0, 1, 2, \ldots\} \]

such that for every \( \omega_1, \omega_2 \in \Omega \) when \( \omega_1 \neq \omega_2 \)
then \( f(\omega_1) \neq f(\omega_2) \)

Means that we can without loss of generality symbolically refer to elements of \( \Omega \) with indices

\( \Omega = \{\omega_1, \omega_2, \ldots, \omega_n\} \) or \( \Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\} \)
Discrete Probability Distribution (2)

Then we define the probability mass function

\[ f(\omega_i) = P(\omega_i) \]

i.e., the probability that an experiment outcome is \( \omega_i \) (or the probability of a random variable having value \( \omega_i \))

The cumulative distribution function is

\[ F(\omega_i) = P(\omega \leq \omega_i) = \sum_{\omega \leq \omega_i} f(\omega) \]

e.g., the enumeration implies an order, this is the probability of the event
Example Uniform distribution on \{1,2,\ldots,n\}

\[ \Omega = \{1,2,\ldots,n\} \]

Every sample equally likely

\[ f(i) = \frac{1}{n} \]

\[ F(j) = j \times \left(\frac{1}{n}\right) = \frac{j}{n} \]
Example: Geometric Distribution

Geometric

- coin flip with probability of H = p
- Number of tosses until first H appears
- $\Omega = \{1, 2, 3, \ldots \}$
- $f(k) = (1-p)^{k-1}p$ assuming independent tosses
- $F(k) = \sum_{1 \leq i \leq k} f(i) =$
Example: Bernoulli Distribution

Experiment: Toss a coin \( n \) times, \( \Pr(H) = p \)

Events of interest : (number of times \( H \) appears)

There are \( 2^n \) possible outcomes (sequence of \( H/T \))

- Probability of sequence is product of toss outcome probabilities, e.g. \( P(HHTTH) = p*p*(1-p)*(1-p)*p \)

Probability that exactly \( j \) head appear is \( B(j,n)*p^j*(1-p)^{n-j} \)

where \( B(n,j) \) is the number of \( n \)-length sequences with exactly \( j \) heads ( “\( n \) choose \( j \)”)

\[
B(n, j) = \binom{n}{j} = \frac{n!}{(n - j)!j!}
\]
Example: Bernoulli Distribution

\[ f(k) = \binom{n}{k} p^k (1 - p)^{n-k} \]

\[ F(k) = \sum_{j=0}^{k} \binom{n}{j} p^j (1 - p)^{n-j} \]
Other functions and metrics

Mean of a distribution (or random variable)

– A single measure that describes a central tendency

\[ \mu = \sum_{j=1}^{\infty} j \times f(j) \quad \text{Or E}[X] \]

\[ \sum_{j=1}^{\infty} j^k \times f(j) \]

K\textsuperscript{th} moment

K\textsuperscript{th} central moment

\[ \sum_{j=1}^{\infty} (j - \mu)^k \times f(j) \quad \text{Or var}(X) \]

2\textsuperscript{nd} central moment is called variance

– standard measure of dispersion
Example comparing dispersions
Computational fun fact

We’ll often make use of the identity

\[ \text{var}(X) = E[X^2] - E[X]^2 \]

Exercise for reader: show that this is true
Hazard rate function

Our analyses will often ask a question like “what is the probability that $X=j$ given that $j \leq X$ ?

– Sometimes called the ‘Hazard rate function’

We have the machinery already to express this

Consider hazard rate function for geometric:
Continuous Probability Distributions

A random variable from a *continuous* distribution takes its values from a non-enumerable set

– Think a range of real numbers

Examples

– Uniform from \([0,1]\) (*any* value \(x\), with \(0 \leq x \leq 1\))
– Time between successive arrivals to a queue
– Amount of service allocated to a job in queue

• Technically an event \(X=t\) cannot have probability
  – Events are built on \(X\) being within a range
Probability Density Function

Probabilities are built up from a distributions probability density function \( f(t) \)

Properties

• \( 0 \leq f(t) \) for all \( t \)

• \( P(s \leq X \leq t) = \int_{s}^{t} f(x) \, dx \)

• Cumulative Distribution Function:

\[
P(X \leq t) = \int_{-\infty}^{t} f(x) \, dx
\]

• \( P(X \leq t) = \int_{-\infty}^{\infty} f(x) \, dx = 1 \)
Common Examples

Exponential
\[ f(t) = \lambda \exp(-\lambda t) \quad \text{for } \lambda > 0 \]
\[ F(t) = 1 - \exp(-\lambda t) \]

Uniform on [0,1]
\[ f(t) = 1 \]
\[ F(t) = t \]

Beta (on [0,1])
\[ f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1 - x)^{\beta-1} \]
Moments
Defined similarly to discrete distributions

\[ \mu = E[X] = \int_{-\infty}^{\infty} x f(x) \, dx \]

\[ E[X^k] = \int_{-\infty}^{\infty} x^k f(x) \, dx \]

\[ \text{var}(X) = \int_{-\infty}^{\infty} (x - E[X])^2 f(x) \, dx \]
Hazard Rate Function

Similar to discrete: \( h(t) = \frac{f(t)}{1 - F(t)} \)

Example: exponential

\[
\frac{f(t)}{1 - F(t)} = \frac{\lambda \exp(-\lambda t)}{1 - (1 - \exp(-\lambda t))} = \lambda
\]

Note independence of hazard rate on \( t \)

• Manifestation of memoryless property
Gaussian Distribution

Gaussian Distributions arise frequently

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

Parameters \( \mu \) (mean) and \( \sigma^2 \) (variance)

\( \mu = 0 \) and \( \sigma^2 = 1 \) is called Standard Normal
Gaussian Distribution

99.7% of the data are within 3 standard deviations of the mean

95% within 2 standard deviations

68% within 1 standard deviation

μ ± 3σ μ ± 2σ μ ± σ μ ± σ μ ± σ μ ± 2σ μ ± 3σ