Review of Probability

Probability

- Random experiment or trial:
  the outcome is not certain, i.e., not known apriority.
- Outcome: the result of a single performance of the random experiment or trial.
  – Continuous or discrete value

  Example:
  – Roll a die and observe number on the top face.
  – Capture a bunny and record the body weight, length of ears, body temperature …

Sample Space

- Sample space S: all possible outcomes

  Example: Rolling a die, the outcome can be 1, 2, …, 6.

  The sample space S ={1, 2, 3, 4, 5, 6}

Event

- The question we might ask: what are the odds of ……?
- A collection of outcomes with common characteristic

  Example 2: Rolling a die and looking for even outcomes define an event E
- Algebra of events:
  – union, intersection, complement

Definition of Probability

- We use the **relative frequency approach**

$$p(A) = \lim_{N\to\infty} \frac{N_A}{N}$$

Axioms of Probability

- The probability p(A) of an event A is a nonnegative fraction in the range

$$0 \leq p(A) \leq 1$$

Axioms of Probability

- The probability of the null event is zero.

$$p(\quad) = 0$$

Axioms of Probability

- The probability of all possible events S is unity

$$p(S) = 1$$

Other Probability Relations

- A and B are two events, then probability of A or B occurring is
  p(A+B) = p(A) + p(B) – p(AB)

Axioms of Probability

- If A and B are **mutually exclusive** events (can not happen at the same time), then the probability that event A **or** event B occurs is
  p(A+B) = p(A) + p(B)

Conditional Probability

- Conditional Probability
  p(A|B) = p(AB) / p(B)
  p(B|A) = p(AB) / p(A)
- For two events, A and B, the probability of both occurring is:
  p(AB) = p(A) p(B|A) = p(B) p(A|B)

- Example: In Europe, 88% of all households have a television. 51% of all households have a television and a DVD. What is the probability that a household has a DVD given that it has a television?

Independent events

- Two events A and B are independent if and only if p(AB) = p(A) p(B)

If two events A and B are independent, then the *conditional probability* of A given B is the same as the "unconditional" (or "marginal") probability of A, that is, p(A|B) = p(A) (and p(B|A) = p(B))

- More generally, any collection of events are **mutually independent** if and only if for any finite subset $A_1, \ldots, A_n$ of the collection we have

$$p(A_1 A_2 \ldots A_n) = p(A_1)\, p(A_2) \ldots p(A_n)$$

- Example 1: Flip a fair coin twice, what is the probability of obtaining two heads?
- Example 2: Four cards are chosen from a standard deck of 52 playing cards **with** replacement. What is the probability of choosing 4 hearts in a row?
- Example 3: Four cards are chosen from a standard deck of 52 playing cards **without** replacement. What is the probability of choosing 4 hearts in a row?
- Example 4: MIT black-jack team hacking Las Vegas by card-counting. *How?*

Partition and Bayes' Theorem

- If $E_1, E_2, \ldots, E_k$ partition the sample space S,

$$p(A) = p(A|E_1)p(E_1)+p(A|E_2)p(E_2)+\ldots+p(A|E_k)p(E_k)$$

- Bayer's theorem: If $\{E_i\}$ forms a partition of S,

$$p(E_i|B) = p(B|E_i)\, p(E_i) \,/\, [\ \textstyle\sum_j p(B|E_j)\, p(E_j)\ ]$$

Example: 1% of baby boys and 2% of baby girls are under-weight. Given a baby is under-weight, what is the chance that this is a boy?

Example: A test can identify a disease as testing positive 99%, and identify non-disease as testing negative 99%. Only 0.001% testers have the disease. Given a positive testing result, what is the chance that the tester has the disease?

Monty Hall Paradox

- Suppose you are on a game show, and you're given the choice of three doors: behind one door is a prize (e.g., a car); behind the others is peanuts. You pick a door, and the host, who knows what's behind the doors, opens another door which has a peanut. He then ask you, "Do you want to switch to the other unopened door?" In this situation, should you switch or not?
- If you are in the "who wants to be a Millionaire" show, and you **have no idea** on a question. You pick one answer randomly initially, and use your 50:50 lifeline. Then, two incorrect answers are removed, but your initially chosen one is still there. Should you switch to the other answer?

Light reading:

A man walked into a Target outside Minneapolis and demanded to see the manager. He was clutching coupons that had been sent to his daughter, and he was angry. "My daughter got this in the mail!" he said. "She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?" The manager looked at the mailer. Sure enough, it was addressed to the man's daughter and contained advertisements for maternity clothing, nursery furniture. The manager apologized and then called a few days later to apologize again.

On the phone, though, the father was somewhat abashed. "I had a talk with my daughter," he said. "It turns out there's been some activities in my house I haven't been completely aware of. She's due in August. I owe you an apology."

http://www.businessinsider.com/the-incredible-story-of-how-target-exposed-a-teen-girls-pregnancy-2012-2

Random Variables

The steps leading to assigning a numerical value to the outcome of a random experiments.

Random Variables

- **Discrete RV**: packet size
- **Continuous RV**: packet delay

Cumulative Distribution Function

- **Probability** that the random variable is less than or equal to x. Thus the event of interest is X ≤ x, and we can write

$$F_X(x) = p(X \leq x)$$

CDF Properties

$$F(-\infty) = 0$$

$$F(\infty) = 1$$

$$0 \leq F(x) \leq 1$$

$$F(x_1) \leq F(x_2) \qquad \text{when } x_1 \leq x_2$$

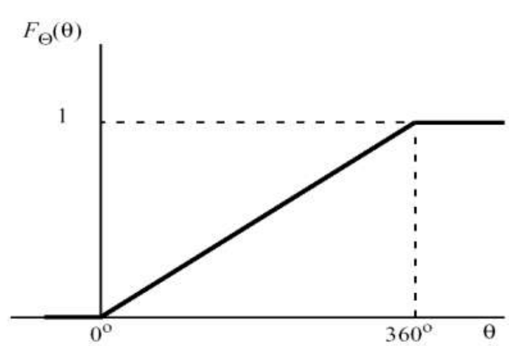
$$p(x_1 < X \leq x_2) = F(x_2) - F(x_1)$$

CDF Properties (cont'd)

- Probability that x lies in the region $x_0 < x \leq x_0 + \epsilon$:

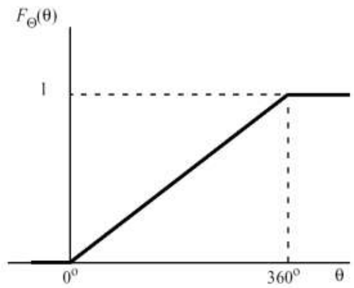
$$p(x_0 < X \leq x_0 + \epsilon) = F(x_0 + \epsilon) - F(x_0)$$

CDF



Cumulative distribution function for a continuous random variable

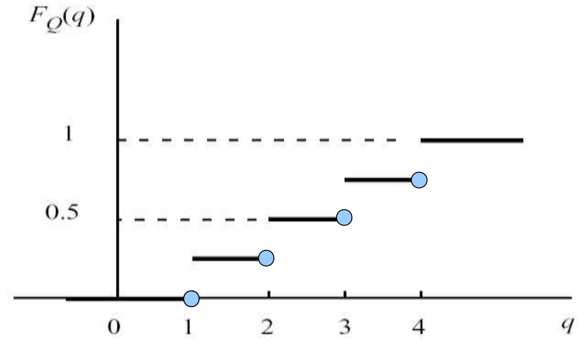
Example: We drop a pen on a table, and the angle of the pen and one edge of the table is a continuous R.V. Θ with the following CDF function:



Find out the probability that $\Theta \leq 180^0$.

Find out the probability that $30^0 < \Theta \leq 120^0$.

CDF



Cumulative distribution function for a discrete random variable.

Probability Density Function (PDF)

$$f_X(x) = \frac{dF_X(x)}{dx}$$

-> For continuous random variables

Probability Density Function (PDF)

- We can also write:

$$\begin{aligned} f_X(x)\,dx &= F_X(x+\delta x) - F_X(x \qquad) \\ &= p(x < X < x+\delta x) \end{aligned}$$

- PDF properties:

$$f_X(x) \geq 0$$
$$\int_{-\infty}^{\infty} f_X(x)dx = F_X(\infty) - F_X(-\infty) = 1$$

Question: can $f_X(x) > 1$?

Example: $f(x) = 2x$ for $0 < x \leq 1$, and $f(x) = 0$ otherwise.

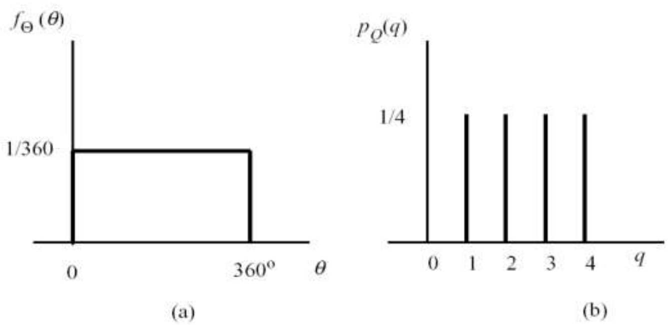
1) can f(x) be a PDF?

2) Find $F_X(x)$.

Probability Mass Function (PMF)

$$p_X(x) \equiv p(X = x)$$

-> For discrete random variables



Continuous and discrete random variables
(a) PDF for the continuous case
(b) PMF for the discrete case.

Expected Value

Continuous RV:

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

Discrete RV:

$$E[X] = \sum_i x_i\, p(x_i)$$

Example:

In a roulette game, bet $1 for any number out of 38, win $35 if the outcome is correctly bet. What is the expected return?

Answer: Let X be the return.

E[X] = (-1) (37/38) + 35 (1/38) = - 1/19

Note: X may not take the value of E[X] which has statistical meaning.

Law of large number: the average of the results obtained from a large number of trials should be close to the expected value.

Note:

E[ ] is a linear operator, similar to + - ∑ ∫

E[X+Y]=E[X]+E[Y]

E[aX]=aE[X], where a is a constant

But, E[XY] may not equal E[X] E[Y].

E[1/X] may not equal 1/E[X].

(see an example: http://panlab.cs.uvic.ca/blog/2011/01/17/dollar-

Variance

$$\sigma^2 = E\left[(X-\mu)^2\right]$$

Independent Random Variables

- Two random variables X and Y say, are said to be independent if and only if the value of X has no influence on the value of Y and vice versa.

More on Expected Value

- $E[X] = E[E[X|Y]]$
  $= E[X|Y=y_1]p(y_1)+E[X|Y=y_2]p(y_2)+ \ldots$
- Example: roll a die, and if the face value X is larger than 3, you can win X dollars; and if X is smaller or equal to 3, you lose 3X dollars.
  - What is the probability of win?
  - What is the expected value of win or lose?

Solution:

1) A: event of win; X_i event of X=i, i=1, 2, ...6

$p(A) = p(A|X_1)p(X_1)+p(A|X_2)p(X_2)+ \ldots$
$+ p(A|X_6)p(X_6) = 0.5$

2) Y: R.V. of the value win/lose

$E[Y] = E[Y|X_1]p(X_1)+E[Y|X_2]p(X_2)+ \ldots +E[Y|X_6]p(X_6)$
$= (-3/6) + (-6/6) + (-9/6) +(4/6)+(5/6)+(6/6)$
$= -0.5$

Will you participate in this game?

Independent Random Variables

- For continuous independent random variables, their probability density functions (PDF) are related by

$$f(x,y) = g(x).h(y)$$

where g(x) and h(y) are the marginal density functions of the random variables X and Y respectively, for all pairs (x,y).

Independent Random Variables

- For discrete independent random variables, their probabilities are related by

$$p(X = x ; Y = y) = p(X = x) p(Y=y)$$

for each pair (x,y).

Conditional Probability

- For continuous random variables, the conditional probability density function

$$f(Y|X) = f(Y, X) / f(X)$$

  - If X and Y are independent
    $f(Y|X) = f(Y)$; $f(X|Y) = f(X)$

Conditional Probability

- For discrete random variables, the conditional probability

$$p(Y | X) = p(XY) / p(X)$$

  - If X and Y are independent
    $p(Y | X) = p(Y)$; $p(X | Y) = p(X)$

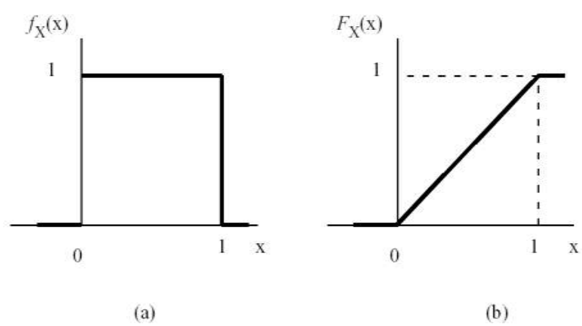
Common RV’s

- Continuous distributions
- Discrete distributions

Continuous Random Variables

Continuous Uniform (Flat) RV

$$f(x) = \begin{cases} 1/(b-a) & a \le x < b \\ 0 & \text{otherwise} \end{cases}$$

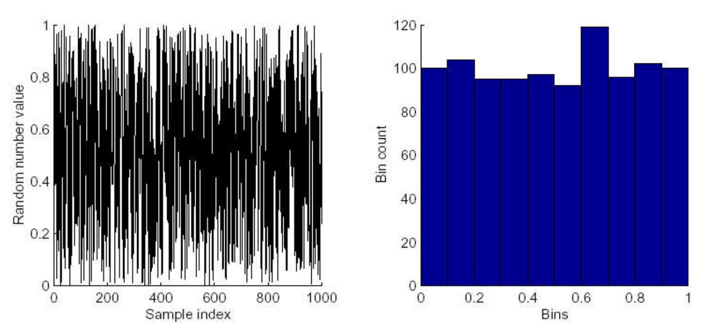


The uniform distribution for a continuous RV:
(a) The PDF; and (b) is the corresponding CDF.

- The mean and variance are

$$\mu = \frac{b+a}{2}$$
$$\sigma^2 = \frac{(b-a)^2}{12}$$

Continuous Uniform (Flat) RV



One thousand samples of a random variable having the uniform distribution in the range 0 to 1 (Left). Histogram for the samples showing a uniform distribution (right)

Exponential RV

$$f(x) = be^{-b\,x} \qquad x \geq 0, b > 0$$

The mean and variance are

$$\begin{aligned} \mu &= \frac{1}{b} \\ \sigma^2 &= \frac{1}{b^2} \end{aligned}$$

Remark: In real world scenarios, the assumption of a constant rate (e.g., arrivals per unit time) is rarely satisfied. For example, the rate of incoming phone calls differs according to the time of day. But if we focus on a time interval during which the rate is roughly constant, such as from 2 to 4 PM during work days, the exponential distribution can be used as a good approximate model for the time until the next phone call arrives. Similar caveats apply to the following examples which yield approximately exponentially distributed variables:

- the time until next earthquake;
- lifetime of a bulb;
- in telecom: the time elapse till the next call arrives;
- in communication networks: the time elapse till the next request to a server;

...

Example:

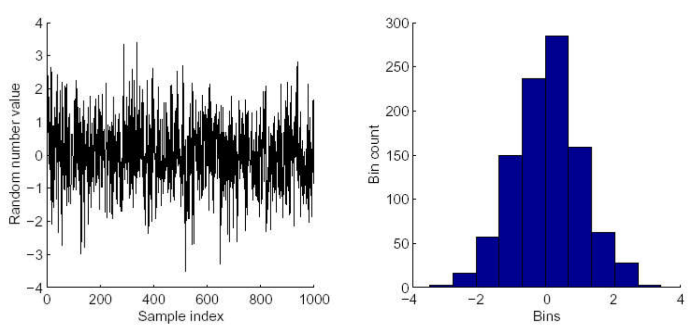
If packets will arrive the router randomly and the inter-arrival time satisfying an exp. distribution with average arrival rate is 1 packets/ms, at time t_0, a packet arrives the router, what is the probability that no packet arrives the router during the following 5 ms?

Given that no packet arrives in 5 ms, what is the probability that no packet arrives for another 5 ms?

Gaussian Random Variable (Normal Distribution)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/(2\sigma^2)}$$

Gaussian RV



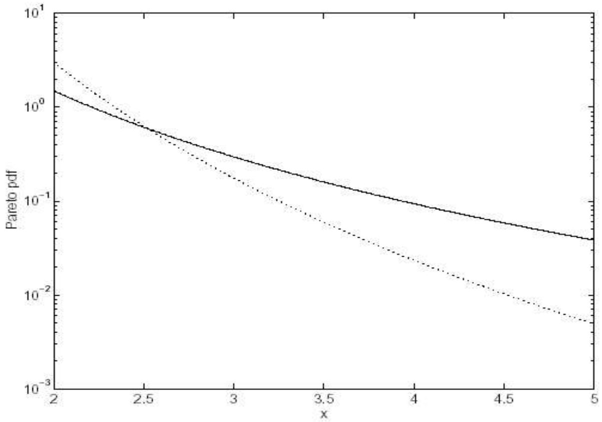
Random numbers generated using the Gaussian distribution with zero mean and unity variance. Figure on left shows the random samples and figure on the right shows their histogram.

Pareto RV

$$f(x) = \frac{b\,a^b}{x^{b+1}}$$

where x ≥ a,

a is the position parameter, b>0 is the shape parameter.



Pareto PDF distribution: a=2 and b=3 (solid line) and b=5 (dashed line).

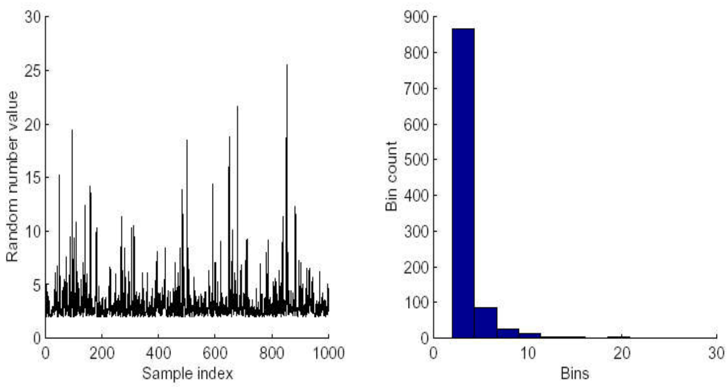
Pareto RV

The mean and variance are

$$\mu = \frac{ba}{b-1}$$

$$\sigma^2 = \frac{b\,a^2}{(b-1)^2\,(b-2)}$$

Pareto RV



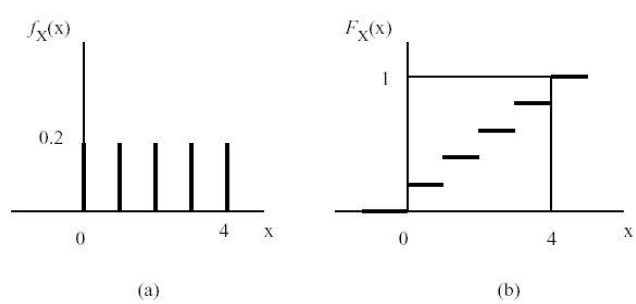
Random numbers generated using the Pareto distribution with a=2 and b=2.5. The figure on left shows the random samples and the figure on the right shows their histogram.

Discrete Random Variables

Discrete Uniform RV

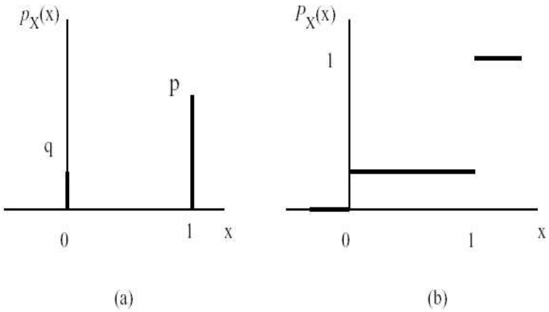
$$p(n) = \begin{cases} 1/k & 1 \le n \le k \\ 0 & \text{otherwise} \end{cases}$$

Discrete Uniform RV



The uniform distribution for a discrete random variable whose values are 0, 1, 2, 3, 4. (a) The PMF; and (b) is the corresponding CDF.

Bernoulli (Binary) RV



(a) PMF; (b) CDF

Bernoulli (Binary) RV

$$p(1) = p$$
$$p(0) = q = 1 - p$$

The mean and variance for are

$$\mu = p$$
$$\sigma^2 = p\,(1-p)$$

Geometric RV

$$p(N = n) = p\,q^n \qquad \text{for } n \geq 0$$

The mean and variance for are

$$\mu = \frac{q}{p}$$

$$\sigma^2 = \frac{q}{p^2}$$

Binomial RV

$$p(K = k) = \binom{N}{k} p^k\, q^{N-k}$$

$$0 \leq k \leq N \text{ and } q = 1 - p$$

The mean and variance for are

$$\mu = N\,p$$
$$\sigma^2 = N\,p\,q$$

DeMoivre-Laplace Theorem

When N >> 1, and Npq >> 1, we can write

$$p(k) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-(k-\mu)^2/2\sigma^2}$$

with

$$\mu = N\,p$$
$$\sigma = \sqrt{N\,p\,q}$$

http://en.wikipedia.org/wiki/File:De_moivre-laplace.gif

- **Example:** A file is being downloaded from a remote site. It has been estimated that on the average 1% of received packets through the channel are in error. Determine the probability that for 500 received packets, 5 of them are in error, using a) the binomial distribution and b) its Gaussian distribution approximation.

Solution:

The binomial distribution parameters are:
N = 500
p = 0.01
q = 0.99
k = 5

The required probability is given by

$$p(K=5) = \binom{500}{5} (0.01)^5 \, (0.99)^{495}$$
$$= 0.1764$$

Using the Gaussian distribution, the parameters are:

$$\mu = N\,p = 5$$
$$\sigma = \sqrt{N\,p\,q} = 2.2249$$

The required probability is given by:

$$p(K=5) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(k-\mu)^2/2\sigma^2}$$
$$= 0.1793$$

This estimate is very close to the more accurate 0.1764

Central Limit Theorem

Given $S_n=(X_1+X_2+...X_n)/n$, where X_i are independent RVs with finite mean μ_i and variance σ_i^2, if n is sufficiently large, then S_n is a Gaussian RV with mean $(\mu_1+\mu_2+...+\mu_n)/n$ and variance $(\sigma_1^2+\sigma_2^2+...+\sigma_n^2)/n^2$

Poisson Theorem

- When N >>1, p<<1, and Np ≈ 1

$$p(k) \approx \frac{(Np)^k}{k!} e^{-Np}$$

with

$$\mu = Np$$
$$\sigma = \sqrt{Npq} = \sqrt{Np(1-p)}$$

Poisson Random Variable

$$p(K = k) = \frac{a^k \, e^{-a}}{k!}$$

An alternative expression

$$p(k) = \frac{(\lambda t)^k \, e^{-\lambda t}}{k!}$$

The mean and variance of k is:

$$\begin{aligned} \mu &= \lambda \, t \\ \sigma^2 &= \lambda \, t \end{aligned}$$

- **Example:** Packets arrive at a device at an average rate of 500 packets per second. Determine the probability that four packets arrive during 3 ms (assuming Poisson distribution).

Solution: We have = 500 (packet/sec), t = 3 ms, and k = 4.

$$p(4) = \frac{(1.5)^4 \, e^{-1.5}}{4!} = 4.7 \times 10^{-2}$$