UCLA STAT 35

Applied Computational and Interactive Probability

•**Instructor**: Ivo Dinov,

Asst. Prof. In Statistics and Neurology

•**Teaching Assistant:** Anwar Khan

University of California, Los Angeles, Winter 2005

http://www.stat.ucla.edu/~dinov/

Discrete Models

Discrete Random Variables and Probability Distributions

Random Variables

Random Variable

For a given sample space S of some experiment, a *random variable* is any rule that associates a number with each outcome in S.

Bernoulli Random Variable

Any random variable whose only possible values are 0 and 1 is called a *Bernoulli random variable.*

Types of Random Variables

A *discrete* random variable is an rv whose possible values either constitute a finite set or else can listed in an infinite sequence. A random variable is *continuous* if its set of possible values consists of an entire interval on a number line.

Probability Distributions for Discrete Random Variables

Probability Distribution

The *probability distribution* or *probability mass function (pmf)* of a discrete rv is defined for every number x by $p(x) =$ P(all s∈ S : X(s) = x)

Parameter of a Probability Distribution

Suppose that $p(x)$ depends on a quantity that can be assigned any one of a number of possible values, each with different value determining a different probability distribution. Such a quantity is called a *parameter* of the distribution. The collection of all distributions for all different parameters is called a *family* of distributions.

Cumulative Distribution Function

The cumulative distribution function (cdf) $F(x)$ of a discrete rv variable X with pmf $p(x)$ is defined for every number by

$$F(x) = P(X \le x) = \sum_{y: y \le x} p(y)$$

For any number x, $F(x)$ is the probability that the observed value of X will be at most x.

Proposition

For any two numbers a and b with $a \le b$,

$$P(a \le X \le b) = F(b) - F(a-)$$

"a–" represents the largest possible X value that is strictly less than (<) a.

Note: For integers

$$P(a \le X \le b) = F(b) - F(a-1)$$

Probability Distribution for the Random Variable X

A probability distribution for a random variable X:

x	–8	–3	–1	0	1	4	6
$P(X = x)$	0.13	0.15	0.17	0.20	0.15	0.11	0.09

Find

a. $P(X \le 0)$ 0.65

b. $P(-3 \le X \le 1)$ 0.67

Expected Values of Discrete Random Variables

The Expected Value of X

Let X be a discrete rv with set of possible values D and pmf $p(x)$. The *expected value* or *mean value* of X, denoted $E(X)$ or μ_X, is

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Example

In the at least one of each or at most 3 children example, where X ={number of Girls} we have:

X	0	1	2	3
pr(x)	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

$$\mathrm{E}(X) = \sum_x x\,\mathrm{P}(x)$$

$$= 0 \times \frac{1}{8} + 1 \times \frac{5}{8} + 2 \times \frac{1}{8} + 3 \times \frac{1}{8}$$

$$= 1.25$$

Ex. Use the data below to find out the expected number of the number of credit cards that a student will possess.

x = # credit cards

x	$P(x = X)$
0	0.08
1	0.28
2	0.38
3	0.16
4	0.06
5	0.03
6	0.01

$$E(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n$$

$$= 0(.08) + 1(.28) + 2(.38) + 3(.16) + 4(.06) + 5(.03) + 6(.01)$$

$$= 1.97$$

About 2 credit cards

The Expected Value of a Function

If the rv X has the set of possible values D and pmf $p(x)$, then the *expected value* of any function $h(x)$, denoted $E[h(X)]$ or $\mu_{h(X)}$, is

$$E[h(X)] = \sum_D h(x) \cdot p(x)$$

Rules of the Expected Value

$$E(aX + b) = a \cdot E(X) + b$$

This leads to the following:

1. For any constant a,
   $E(aX) = a \cdot E(X)$.
2. For any constant b,
   $E(X + b) = E(X) + b$.

The Variance and Standard Deviation

Let X have pmf $p(x)$, and expected value μ
Then the *variance* of X, denoted $V(X)$ (or σ_X^2 or σ^2), is

$$V(X) = \sum_D (x-\mu)^2 \cdot p(x) = E[(X-\mu)^2]$$

The *standard deviation* (SD) of X is

$$\sigma_X = \sqrt{\sigma_X^2}$$

Ex. The quiz scores for a particular student are given below:

22, 25, 20, 18, 12, 20, 24, 20, 20, 25, 24, 25, 18

Find the variance and standard deviation.

Value	12	18	20	22	24	25
Frequency	1	2	4	1	2	3
Probability	.08	.15	.31	.08	.15	.23

$\mu = 21$

$$V(X) = p_1(x_1-\mu)^2 + p_2(x_2-\mu)^2 + \ldots + p_n(x_n-\mu)^2$$

$$\sigma = \sqrt{V(X)}$$

$$V(X) = .08(12-21)^2 + .15(18-21)^2 + .31(20-21)^2$$
$$+ .08(22-21)^2 + .15(24-21)^2 + .23(25-21)^2$$

$$V(X) = 13.25$$

$$\sigma = \sqrt{V(X)} \quad = \sqrt{13.25} \approx 3.64$$

Shortcut Formula for Variance

$$V(X) = \sigma^2 = \left[\sum_D x^2 \cdot p(x)\right] - \mu^2$$
$$= E\left(X^2\right) - \left[E(X)\right]^2$$

Rules of Variance

$$V(aX+b) = \sigma_{aX+b}^2 = a^2 \cdot \sigma_X^2$$

and $\sigma_{aX+b} = |a| \cdot \sigma_X$

This leads to the following:

1. $\sigma_{aX}^2 = a^2 \cdot \sigma_X^2, \sigma_{aX} = |a| \cdot \sigma_X$
2. $\sigma_{X+b}^2 = \sigma_X^2$

Linear Scaling (affine transformations) *aX* + *b*

For any constants *a* and *b*, the expectation of the RV ***aX* + *b*** is equal to the sum of the product of a and the expectation of the RV *X* and the constant *b*.

E(*a**X*** + *b*) = ***a*** E(***X***) +*b*

And similarly for the standard deviation (*b*, an additive factor, does not affect the SD).

SD(*a**X*** +*b*) = |***a***| SD(***X***)

Linear Scaling (affine transformations) *aX* + *b*

Why is that so?

E(*a**X*** + *b*) = ***a*** E(***X***) +*b*   SD(*a**X*** +*b*) = |***a***| SD(***X***)

$$E(aX+b)=\sum_{x=0}^{n}(a\,x+b)\;P(X=x)=$$

$$\sum_{x=0}^{n}a\,x\,P(X=x)+\sum_{x=0}^{n}b\;P(X=x)=$$

$$a\sum_{x=0}^{n}x\,P(X=x)+b\sum_{x=0}^{n}P(X=x)=$$

$$aE(X)+b\times 1=aE(X)+b.$$

Linear Scaling (affine transformations) *aX* + *b*

Example:

E(*a**X*** + *b*) = ***a*** E(***X***) +*b*   SD(*a**X*** +*b*) = |***a***| SD(***X***)

1. X={-1, 2, 0, 3, 4, 0, -2, 1}; P(X=x)=1/8, for each x
2. Y = 2X-5 = {-7, -1, -5, 1, 3, -5, -9, -3}
3. E(X)=
4. E(Y)=
5. Does **E(X) = 2 E(X) –5 ?**
6. Compute SD(X), SD(Y). **Does SD(Y) = 2 SD(X)?**

Linear Scaling (affine transformations) *aX* + *b*

And why do we care?

E(*a**X*** + *b*) = ***a*** E(***X***) +*b*   SD(*a**X*** +*b*) = |***a***| SD(***X***)

-completely general strategy for computing the distributions of RV's which are obtained from other RV's with known distribution. E.g., X~N(0,1), and Y=aX+b, then we need **not** calculate the mean and the SD of Y. We know from the above formulas that E(Y) = b and SD(Y) =|a|.

-These formulas hold for all distributions, not only for Binomial and Normal.

Linear Scaling (affine transformations) *aX* + *b*

And why do we care?

E(*a**X*** + *b*) = ***a*** E(***X***) +*b*   SD(*a**X*** +*b*) = |***a***| SD(***X***)

-E.g., say the rules for the game of chance we saw before change and the new pay-off is as follows: {$0, $1.50, $3}, with probabilities of {0.6, 0.3, 0.1}, as before. What is the newly expected return of the game? Remember the old expectation was equal to the entrance fee of $1.50, and the game was fair!

Y = 3(X-1)/2

{$1, $2, $3} → {$0, $1.50, $3},

E(Y) = 3/2 E(X) –3/2 = 3 / 4 = $0.75

And the game became clearly biased. Note how easy it is to compute E(Y).

Means and Variances for (in)dependent Variables!

- **Means**:
  - Independent/Dependent Variables {X1, X2, X3, …, X10}
    - E(X1 + X2 + X3 + … + X10) = E(X1)+ E(X2)+ E(X3)+… + E(X10)
- **Variances**:
  - Independent Variables {X1, X2, X3, …, X10}, variances add-up
    Var(X1 +X2 + X3 + … + X10) = Var(X1)+Var(X2)+Var(X3)+…+Var(X1)
  - Dependent Variables {X1, X2}
    Variance contingent on the variable dependences,
    - E.g., If X2 = 2X1 + 5,

Var(X1 +X2) =Var (X1 + 2X1 +5) =

Var(3X1 +5) =Var(3X1) = 9Var(X1)

The Binomial Probability Distribution

Binomial Experiment

An experiment for which the following four conditions are satisfied is called a *binomial experiment*.

1. The experiment consists of a sequence of n trials, where n is fixed in advance of the experiment.

2. The trials are identical, and each trial can result in one of the same two possible outcomes, which are denoted by success (S) or failure (F).
3. The trials are independent.
4. The probability of success is constant from trial to trial: denoted by p.

Binomial Experiment

Suppose each trial of an experiment can result in S or F, but the sampling is without replacement from a population of size N. If the sample size n is at most 5% of the population size, the experiment can be analyzed as though it were exactly a binomial experiment.

Binomial Random Variable

Given a binomial experiment consisting of n trials, the *binomial random variable* X associated with this experiment is defined as

X = the number of S's among n trials

Notation for the pmf of a Binomial rv

Because the pmf of a binomial rv X depends on the two parameters n and p, we denote the pmf by $b(x;n,p)$.

Computation of a Binomial pmf

$$b(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$0 \le x \le n$$

Ex. A card is drawn from a standard 52-card deck. If drawing a club is considered a success, find the probability of

a. exactly one success in 4 draws (with replacement).

$p = ¼; q = 1 – ¼ = ¾$

$$\binom{4}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^3 \approx 0.422$$

b. no successes in 5 draws (with replacement).

$$\binom{5}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^5 \approx 0.237$$

Slide 37 Stat 35, UCLA, Ivo Dinov

Notation for cdf

For $X \sim \text{Bin}(n, p)$, the cdf will be denoted by

$$P(X \le x) = B(x; n, p) = \sum_{y=0}^{x} b(y; n, p)$$

$x = 0, 1, 2, \ldots n$

Slide 38 Stat 35, UCLA, Ivo Dinov

Mean and Variance

For $X \sim \text{Bin}(n, p)$, then $E(X) = np$, $V(X) = np(1 - p) = npq, \sigma_X = \sqrt{npq}$ (where $q = 1 - p$).

Slide 39 Stat 35, UCLA, Ivo Dinov

Ex. 5 cards are drawn, with replacement, from a standard 52-card deck. If drawing a club is considered a success, find the mean, variance, and standard deviation of X (where X is the number of successes).

$p = ¼; q = 1 – ¼ = ¾$

$$\mu = np = 5\left(\frac{1}{4}\right) = 1.25$$

$$V(X) = npq = 5\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = 0.9375$$

$$\sigma_X = \sqrt{npq} = \sqrt{0.9375} \approx 0.968$$

Slide 40 Stat 35, UCLA, Ivo Dinov

Ex. If the probability of a student successfully passing this course (C or better) is 0.82, find the probability that given 8 students

a. all 8 pass. $\binom{8}{8}(0.82)^8(0.18)^0 \approx 0.2044$

b. none pass. $\binom{8}{0}(0.82)^0(0.18)^8 \approx 0.0000011$

c. at least 6 pass.

$$\binom{8}{6}(0.82)^6(0.18)^2 + \binom{8}{7}(0.82)^7(0.18)^1 + \binom{8}{8}(0.82)^8(0.18)^0$$

$$\approx 0.2758 + 0.3590 + 0.2044 = 0.8392$$

Slide 41 Stat 35, UCLA, Ivo Dinov

Hypergeometric and Negative Binomial Distributions

Slide 42 Stat 35, UCLA, Ivo Dinov

The Hypergeometric Distribution

The three assumptions that lead to a *hypergeometric distribution*:

1. The population or set to be sampled consists of N individuals, objects, or elements (a finite population).

2. Each individual can be characterized as a success (S) or failure (F), and there are M successes in the population.

3. A sample of n individuals is selected without replacement in such a way that each subset of size n is equally likely to be chosen.

Hypergeometric Distribution

If X is the number of S's in a completely random sample of size n drawn from a population consisting of M S's and $(N - M)$ F's, then the probability distribution of X, called the hypergeometric distribution, is given by

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$\max(0, n - N + M) \le x \le \min(n, M)$$

Hypergeometric Mean and Variance

$$E(X) = n \cdot \frac{M}{N} \qquad V(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N}\left(1 - \frac{M}{N}\right)$$

The Negative Binomial Distribution

The *negative binomial rv* and *distribution* are based on an experiment satisfying the following four conditions:

1. The experiment consists of a sequence of independent trials.

2. Each trial can result in a success (S) or a failure (F).

3. The probability of success is constant from trial to trial, so $P(S \text{ on trial } i) = p$ for $i = 1, 2, 3, \ldots$

4. The experiment continues until a total of r successes have been observed, where r is a specified positive integer.

pmf of a Negative Binomial

The pmf of the negative binomial rv X with parameters r = number of S's and $p = P(S)$ is

$$NB(x; r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x$$

x = 0, 1, 2, …

Negative Binomial Mean and Variance

$$E(X) = \frac{r(1-p)}{p} \qquad V(X) = \frac{r(1-p)}{p^2}$$

Hypergeometric Distribution & Binomial

- Binomial approximation to Hyperheometric
  - $\frac{n}{N}$ is small (usually < 0.1), then $\frac{M}{N} \approx p$

$$HyperGeom(x; N, n, M) \underset{M/N = p}{\overset{approaches}{\Longrightarrow}} Bin(x; n, p)$$

Ex: 4,000 out of 10,000 residents are against a new tax. 15 residents are selected at random.

P(at most 7 favor the new tax) = ?

Geometric, Hypergeometric, Negative Binomial

- Negative binomial pmf [X ~ NegBin(r, p), if r=1 ➔ Geometric (p)] $P(X = x) = (1-p)^x p$

Number of trials (x) until the r[th] success (negative, since number of successes (r) is fixed & number of trials (X) is random)

$$P(X = x) = \binom{x+r-1}{r-1} p^r (1-p)^x$$

$$E(X) = \frac{r(1-p)}{p}; \quad Var(X) = \frac{r(1-p)}{p^2}$$

The Poisson Probability Distribution

Poisson Distribution

A random variable X is said to have a *Poisson distribution* with parameter λ $(\lambda > 0)$, if the pmf of X is

$$p(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \qquad x = 0, 1, 2...$$

The Poisson Distribution as a Limit

Suppose that in the binomial pmf $b(x;n, p)$, we let $n \to \infty$ and $p \to 0$ in such a way that np approaches a value $\lambda > 0$. Then $b(x; n, p) \to p(x; \lambda)$.

Poisson Distribution Mean and Variance

If X has a Poisson distribution with parameter λ, then

$$E(X) = V(X) = \lambda$$

Poisson Process

3 Assumptions:

1. There exists a parameter $\alpha > 0$ such that for any short time interval of length Δt, the probability that exactly one event is received is $\alpha \cdot \Delta t + o(\Delta t)$.

2. The probability of more than one event during Δt is $o(\Delta t)$.
3. The number of events during the time interval Δt is independent of the number that occurred prior to this time interval.

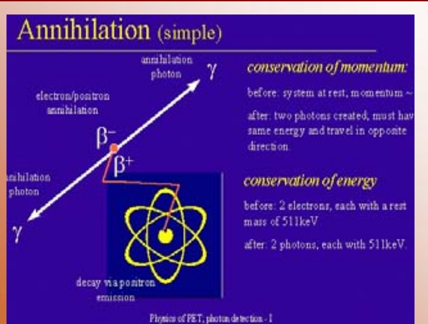
Poisson Distribution

$P_k(t) = e^{-\alpha t} \cdot (\alpha t)^k / k!$, so that the number of pulses (events) during a time interval of length t is a Poisson rv with parameter $\lambda = \alpha t$. The expected number of pulses (events) during any such time interval is αt, so the expected number during a unit time interval is α.

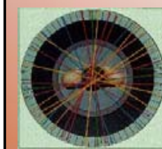
Poisson Distribution – Definition

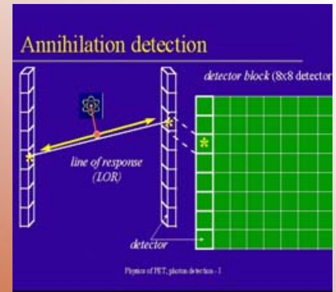
- Used to model counts – number of arrivals (k) on a given interval …
- The Poisson distribution is also sometimes referred to as the **distribution of rare events**. Examples of Poisson distributed variables are number of accidents per person, number of sweepstakes won per person, or the number of catastrophic defects found in a production process.

Functional Brain Imaging – Positron Emission Tomography (PET)



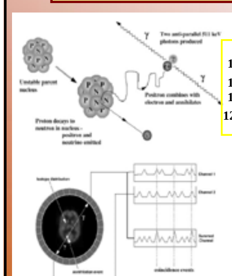
Functional Brain Imaging - Positron Emission Tomography (PET)





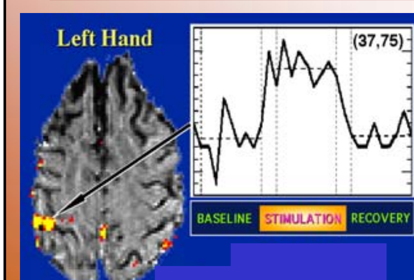
http://www.nucmed.buffalo.edu

Functional Brain Imaging – Positron Emission Tomography (PET)



Isotope	Energy (MeV)	Range(mm)	1/2-life	Appl.
^{11}C	0.96	1.1	20 min	receptors
^{15}O	1.7	1.5	2 min	stroke/activation
^{18}F	0.6	1.0	110 min	neurology
^{124}I	~2.0	1.6	4.5 days	oncology

Functional Brain Imaging – Positron Emission Tomography (PET)



Hypergeometric Distribution & Binomial

- Binomial approximation to Hyperheometric
  - $\frac{n}{N}$ is small (usually < 0.1), then $\frac{M}{N} \approx p$

$$HyperGeom(x; N, n, M) \overset{approaches}{\underset{M/N=p}{\Longrightarrow}} Bin(x; n, p)$$

Ex: 4,000 out of 10,000 residents are against a new tax. 15 residents are selected at random.

P(at most 7 favor the new tax) = ?

Poisson Distribution – Mean

- Used to model counts – number of arrivals (k) on a given interval …
- Y~Poisson(λ), then P(Y=k) = $\frac{\lambda^k e^{-\lambda}}{k!}$, k = 0, 1, 2, …
- Mean of Y, $\mu_Y = \lambda$, since

$$E(Y) = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} =$$

$$= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Poisson Distribution - Variance

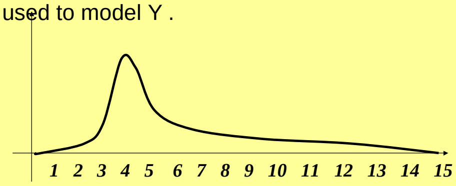
- $Y \sim Poisson(\lambda)$, then $P(Y=k) = \frac{\lambda^k e^{-\lambda}}{k!}$, k = 0, 1, 2, …
- Variance of Y, $\sigma_Y = \lambda^{1/2}$, since

$$\sigma_Y^2 = Var(Y) = \sum_{k=0}^{\infty}(k-\lambda)^2 \frac{\lambda^k e^{-\lambda}}{k!} = \ldots = \lambda$$

- For example, suppose that Y denotes the number of blocked shots (arrivals) in a randomly sampled game for the UCLA Bruins men's basketball team. Then a Poisson distribution with mean=4 *may be* used to model Y .

Poisson Distribution - Example

- For example, suppose that Y denotes the number of blocked shots in a randomly sampled game for the UCLA Bruins men's basketball team. Poisson distribution with mean=4 may be used to model Y .



Poisson as an *approximation* to Binomial

- Suppose we have a sequence of Binomial(n, p_n) models, with $\lim(n\, p_n) \rightarrow \lambda$, as $n \rightarrow$ infinity.
- For each 0<=y<=n, if $Y_n \sim$ Binomial(n, p_n), then
  - $P(Y_n=y) = \binom{n}{y} p_n^{\,y}(1-p_n)^{n-y}$
  - But this converges to:

$$\binom{n}{y} p_n^{\,y}(1-p_n)^{n-y} \xrightarrow[n \to \infty,\ n \times p_n \to \lambda]{\text{WHY?}} \frac{\lambda^y e^{-\lambda}}{y!}$$

- Thus, Binomial(n, p_n) → Poisson(λ)

Poisson as an *approximation* to Binomial

- **Rule of thumb** is that approximation is good if:
  - **n>=100**
  - **p<=0.01**
  - **λ =n p <=20**
- Then, Binomial(n, p_n) → Poisson(λ)

Example using Poisson approx to Binomial

- Suppose P(defective chip) = 0.0001=10^{-4}. Find the probability that a lot of 25,000 chips has > 2 defective!
- Y~ Binomial(25,000, 0.0001), find P(Y>2). Note that Z~Poisson(λ =n p =25,000 x 0.0001=2.5)

$$P(Z>2) = 1 - P(Z \le 2) = 1 - \sum_{z=0}^{2} \frac{2.5^z}{z!} e^{-2.5} =$$

$$1 - \left(\frac{2.5^0}{0!}e^{-2.5} + \frac{2.5^1}{1!}e^{-2.5} + \frac{2.5^2}{2!}e^{-2.5}\right) = 0.456$$

Relation among Distributions