Week 2 - Distributions

Statistics is the study of distributions

Types of variables

- Continuous (real)
  - Temperature, nitrogen concentration, time, steroid concentration
- Discrete
  - Ordered, infinite
    - Number of individuals
  - Unordered, finite
    - Blue morph, green morph
    - Cooperative, uncooperative
- Statistical tools available entirely dependent on types of variables

Tool depends on variables

	Dependent Discrete	Dependent Continuous
Independent Discrete	Contingency table/chi2	ANOVA t-test
Independent Continous	Logistic regression CART	(linear) regression CART

The top 10 things you need to remember from calculus

1. Derivatives=slope – A derivative is the slope of the line tangent to the function dy/dx
2. Differential=rate of change in variables – A differential represents how much y changes with a change in x: dy=df(x)dx. Note that for infinitesimally small dx, this is completely accurate, as dx gets bigger, this becomes less accurate (in fact it is a first order or linear approximation).
3. Derivatives and differentials are defined at a single point, but if we use this definition at every point, we get a new function. This is the derivative of f(x) and is denoted f'(x)

Top 10 (continued)

4. Simple rules for taking derivatives are:

$f(x)=c$	$f'(x)=0$
$f(x)=ax$	$f(x)=a$
$f(x)=x^n$	$f'(x)=nx^{n-1}$
$f(x)=exp(x)$	$f(x)=exp(x)$
$f(x)=ln(x)$	$f(x)=1/x$
$f(x)=g(x)*h(x)$	$f(x)=g'(x)h(x)+g(x)h'(x)$
$f(x)=g(h(x))$	$f'(x)=g'(h(x))h'(x)$
$f(x)=1/g(x)$	$f'(x)=-g'(x)/g(x)^2$

Top 10 (continued)

5. Derivatives can be taken of derivatives. This yields second derivatives and so on. A second derivative tells how fast the slope is changing. A positive second derivative means the slope is increasing and hence is concave up.
6. An integral is an antiderivative
7. An integral is an area under a curve
8. An integral is an infinite sum (uncountable)

Top 10 (continued)

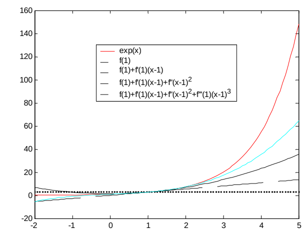
9. The geometric series:

$$(1-r)(1+r+r^2+r^3+...+r^n)=1-r^{n+1} \Rightarrow \sum_{i=m}^{n} a^i = \frac{a^m(1-r^{n+1})}{1-r} = \frac{a^m}{1-r} \; if\,|r|<1, n=\infty$$

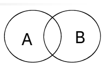
10. The Taylor expansion:

$$f(x)=\sum_{n=0}^{\infty}\frac{f^n(a)}{n!}(x-a)^n \approx f(a)+f'(a)(x-a)+f''(a)(x-a)^2+...$$

Taylor's series



Probability



1) Law of simultaneous events:
- $P(\text{both A and B occur})=P(A\cup B)= P(A)+P(B)-P(A\cap B)$

1a) Law of addition
- a special case of above: if $A\cap B=\varnothing \Rightarrow P(A\cup B)= P(A)+P(B)$

2) Conditional probability
- $P(A|B)=P(A\cap B)/P(B)$

3) Independence
- If $P(A|B)=P(A)$ then A & B are independent

23a) Law of multiplication
- combining 2 & 3
- if A & B are independent then $P(A \cap B)=P(A)P(B)$

4) Law of total probability
- Let Ai be a disjoint set that spans Σ (i.e. $A_i \cap A_j = \varnothing$ and $\cup_i A_i = \Sigma$)
- Then for any B, $P(B)=\Sigma P(B\cap A_i)=\Sigma P(B|A_i)P(A_i)$

Concrete example

- Roll a die
  - A={1,3,5} Odd numbers (3/6)
  - B={1,2,3} Low numbers (3/6)
  - C={6} Highest number (1/6)
  - D={1,2} Low numbers (2/6)
- Law of simultaneous events:
  - $P(A \cup B) = P(A)+P(B)-P(A \cap B)$
    - =3/6+3/6-2/6=4/6
- Law of addition
  - $P(A \cup C)=P(A)+P(C)=3/6+1/6=4/6$ ($P(A \cap B)=0$)
- Independence
  - $P(A|B)=P(A \cap B)/P(B) = 2/6 / 3/6 = 2/3 \neq P(A)$
  - $P(A|D)= P(A \cap D)/P(D) = 1/6 / 2/6 = ½ = P(A)$
- Law of multiplication
  - $P(A \cap D)=P(A)*P(D)=3/6*2/6=1/6$
- Law of total probability
  - P(Low)=P(Low|Even)*P(Even)+P(Low|Odd)*P(Odd)
  - P(B)=P(B|~A)*P(~A)+P(B|A)*P(A)=1/3*1/2+2/3*1/2=1/2

Random variables

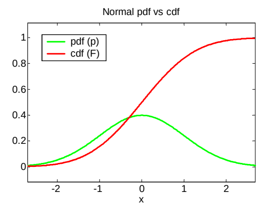
- A random variable (X) takes on different values
  - may be continuous or discrete
  - May go form $-\infty$ to $+\infty$ or just 0 to $+\infty$ or ...
- A random variable is defined by its distribution
  - E.g. X~N(0,1)
- The probability that it takes on a given value is completely specified by a function:
  - $F(x)=P(X\leq x)$
  - E.g. if X~N(0,1) then F is hard to write down but know the picture

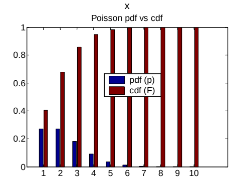
Two cases

- {x} discrete (1,2,3,4,...)
  - Usually specify p(x)=P(X=x)
  - Then $F(x) = P(X \le x) = \sum_{i=1}^{x} p(x)$
- {x} continuous
  - Sometimes easier to specify p(x)
    - No exact meaning (probability N(0,1)=0.134278=0)
  - But $F(x) = P(X \le x) = \int_{-\infty}^{x} p(x)dx$
  - E.g. Normal $p(x)=1/\mathrm{sqrt}(2\pi)\sigma*\exp(-(x-\mu)^2/\ \sigma^2)$
- i.e. Density vs cumulative probability function

Notation alert

- Throughout I will use:
  - P=probability of
  - F = cumulative distribution function
  - p = probability density
  - F=∫p





Expectation

- Expectation is a fancy word for a weighted average
  - Specifically averaged weighted by p(x)

$$E(x) = \sum_i x_i p(x_i) = \int_{-\infty}^{\infty} xp(x)dx$$

  - Gives the mean (μ=E(X))
    - $(x_1+x_2+x_3)/n = 1/n\ x_1+1/n\ x_2+ 1/n\ x_3$
- Example
  - Empirical distribution 1,1,3,3,4
  - 1*.4+3*.4+4

Moments are a generalization of expectation

- i^{th} moment is μ_i=E(X^n)

$$E(x^n) = \sum_i x_i^n p(x_i) = \int_{-\infty}^{\infty} x^n p(x)dx$$

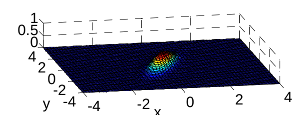
- Generally more useful is a "central moment)
  - μ_i=E((X-E(X))^n)
- The 2^{nd} central moment is the variance
  - 3^{rd} is skew, 4^{th} is kurtosis
- Trick:
  - $Var(X)=E((X-\mu)^2)=E(X^2-2\mu X+\mu^2)=$
  - $E(X^2)-2\mu E(X)+\mu^2 = E(X^2)-E(X)^2$
- Beware E(f(X)) ≠ f(E(X))
  - Exception E(a+bX)=a+bE(X)

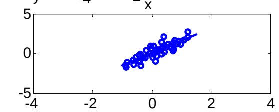
Bivariate distributions

- So far univariate: p(x) and P(X ≤ x)
  - Bivariate p(x,y), P(X ≤ x & Y ≤ y)
    - Regression, correlation derive from this view
  - Multivariate – n-way
    - Useful conceptualization of multivariate data

$$F(x,y) = P(X \le x, Y \le y) = \sum_{i=1,j=1}^{i=x,j=y} p(x,y)$$

Bivariate distributions





σ_1=0.5, σ_2=1.0, ρ=0.8 ($\sigma_{12}=\rho*\sigma_1*\sigma_2$=0.4)

Bivariate expectations

- Conditional expectation
  - $E(Y|X)$ is well defined
  - Average value of X if we know Y happened
  - Analog with regression $E(Y|X)$ or $y=a+bx$
- Covariance
  - $Cov(X,Y)=E((X-\mu_x)(Y-\mu_y)) = 1/n \, \Sigma \, (x_i-\mu_x)(y_i-\mu_y)$
    - Shortcut calculation: $Cov(X,Y)=E(XY)-E(X)E(Y)$
    - Note: $Var(X)=Cov(X,X)$
    - $\sigma_{xy}^2=Cov(X,Y)$
  - $Cov(X,Y)=0 \Leftrightarrow$ X & Y independent
- General regression/correlation
  - $\rho=\sigma_{xy}/\sigma_x\sigma_y$
  - $b=\sigma_{xy}/\sigma_x^2$

Some important limits

- Law of large numbers
  - Sample average → population average
  - $1/n\Sigma x_i = \mu$ as $n \to \infty$
  - Samples must be independent
- More generally,
  - Sample distribution → population
- Central limit theorem
  - Take any random variable (continuous, discrete)
  - Take a large sample (independent!)
  - Sum together
  - →Normal
  - $\Sigma X \to \sim N(n\mu,\sqrt{n}\sigma)$ for any X

The 13 most important distribution functions

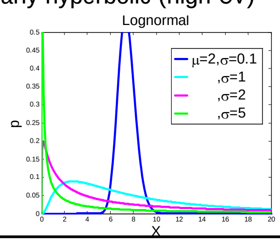
- Normal group (5)

Combination	**Distribution**	**Range**
Sample mean from n normals $N(\mu,\sigma)$	$N(\mu,\sigma/\sqrt{n})$	$(-\infty,\infty)$
Sum n normals squared: N(0,1)	Chi-squared χ_n	$[0,\infty)$
$N(0,1)/\sqrt{\chi_n}$ (found in sample distribution of variance)	t-distribution	$(-\infty,\infty)$
χ_n/χ_m (found in ANOVA tests)	Fisher	$[0,\infty)$
log N(0,1) (found as central limit of *product* of many variables)	Log-normal	$(0,\infty)$

- Recall t→Normal as n large

Lornormal

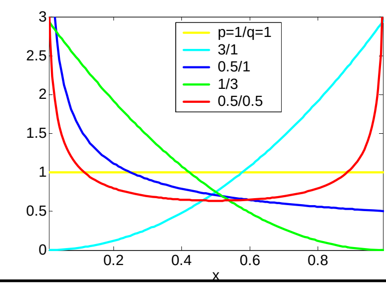
- Ranges from nearly normal (low CV) to nearly hyperbolic (high CV)



Finite range group (2)

- Uniform
- Beta is a generalization
  - Two parameters
  - $p(x)=Cx^{p-1}(1-x)^{q-1}$
    - C is a constant such that sum(p)=1
  - p controls behavior left side (p>1, then like x^2, p<1→hyperbolic, p=1 → horizontal)
  - q controls right side
- Uniform is p=q=1
- Beta is solution of distribution of allele frequencies under drift (+selection)

Beta distribution



Counting process distributions (6)

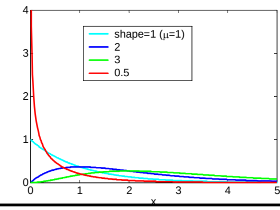
- Imagine you are sitting some where
  - Events occur
  - You study the # of events, the time between events
- This is a counting process
- Now time can be:
  - discrete (e.g. did the event occur this year)
  - continuous (e.g. Geiger counter)
- These all assume events at time t are independent of all other times
  - Memoryless

6 Counting process distributions

	Discrete Time			Continuous Time		
	Time=n, rate at single interval=p			Time=t, instantaneous rate=λ		
	Distribution	**E(X)**	**Var(X)**	**Distribution**	**E(X)**	**Var(X)**
time to next event	Geometric	$1/p$	$(1-p)/p^2$	Exponential	$1/\lambda$	$1/\lambda^2$
time to the nth event	Negative binomial	n/p	$n(1-p)/p^2$	Gamma	n/λ	n/λ^2
expected (mean) # events in given time	Binomial	np	$np(1-p)$	Poisson	λt	λt

Gamma

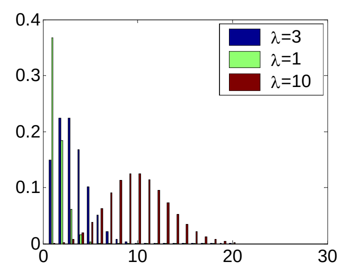
- Gamma is highly plastic and therefore useful for functions on interval $(0,\infty)$



Poisson

- Law of rare events
  - Like central limit theorem
  - As long as events independent →Poisson
  - A good model of counted numbers
  - E.g. deaths by horse kick in the Prussian army
  - If your data are counts think Poisson
- Note E(X) for Poisson is λt
- Poisson in space as well!
  - 1,2, 3 dimensions
  - Random placement in space
  - $N=\lambda A$

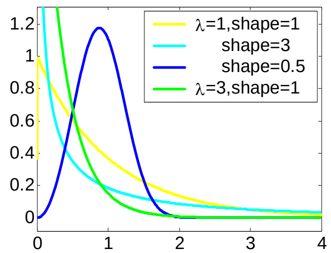
Poisson



Generalizations

- Binomial
  - Hypergeometric (finite # of successes possible – i.e. no independence of events)
  - Multinomial (more than 2 outcomes)
- Exponential
  - Power or Pareto or 1/f or hyperbolic
    - X^{-c} instead of c^{-x} for exponential
  - Weibull – wait times with memory
    - Exponetial+Pareto
    - Shape parameter
      - 1=exponential
      - >1=recur quickly or wait a long time
      - <1= consistent waiting interval

Weibull



Summary – 16 distributions

- Normal group (5)
  - Normal, Lognormal & stats (t, F, χ^2)
- Finite interval (2)
  - Uniform $\subseteq$ Beta
- Counting processes (x6)
  - Time to 1 event, time to n events, # events in time t
  - Binary/discrete
  - Exponential $\subseteq$ Gamma, Poisson
- Counting process generalizations (3)
  - Binomial-> hypergeometric, multinomial
  - Exponential->Weibull, Pareto

What distribution?

- Time between sightings of rare bird
- # of rare birds seen in x hours
- Time between fires
- # of tent caterpillar nests in a tree
- Time between encounters with prey items
- Abundances over time in stochastic environment
- Allele frequencies
- Body sizes
- # of years with killing frost in November

Arithmetic on distributions

Combining distributions

- If know X~?, Y~?, what is:
  - X+Y~
  - X*Y~
- Why care:
  - Often encountered in data analysis
  - Mean food intake =?? Mean prey weight * mean kcal/g
  - Know N~ , M~, what is distribution of reproductive effort (N*M)
  - Life history:

    E(Time to reproduction) =?? E(germination time) + E(maturation time)

Practical example

N	M	RE
4	1	4
3	2	6
2	3	6
2	2	4
1	5	5

- E(N)=12/5=2.4
- E(M)=13/5=2.6
- E(RE)=25/5=5
- E(M)*E(N) =6.24 ≠ 5=E(RE)
- Would be true if N, M indepdenent
  - But in this example there is a well-known inverse tradeoff – NEVER independent

Combining distributions - answer

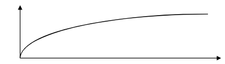
- Generally a hard question
  - Uses “generating functions”
  - Some well known:
    - Sum of N(a,b)+N(c,b)=N(a+c,b)
    - Sum $\Sigma_n N(a,b) \sim N(n*a, b*n^{-1/2})$
    - Sum exponentials and/or gamma is gamma
    - Sum of Poisson is poisson
  - Generally hard to solve
  - Evans, Hasting & Peacock a good summary

Can we at least answer for moments?

- E(a+bX+cY)=a+bE(X)+cE(Y)
- E(XY)=E(X)E(Y)+Cov(X,Y)
  - Simple case only if X,Y independent
- $E(1/X) \approx 1/E(X)+Var(X)/E(X)^3$
- E(X/Y) various (See Welch et al)
- $Var(aX)=a^2Var(X)$
- Var(X+Y)=var(X)+var(Y)+2*Cov(X,Y)
  - Simple case only if X,Y independent

Distribution of g(X)

- If know distribution of random variable: X~?, what is distribution of g(X)
- Why care:
  - X is prey density, g(X) is intake rate
  - X is age of reproduction, g(X) is fitness
  - X is nitrogen concentration, g(X) is yield

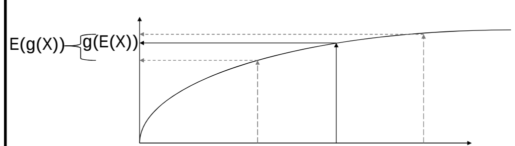


Solving g(X)~?

- Few cases (mostly if X normal)
  - If X is normal, exp(X) is lognormal
  - If X is normal, X^2 is Chi-squared
- General solution hard mathematically:
  - Let p be pdf of X, f be pdf of g(X)
  - Then $f(x)=p(g^{-1}(y))\ dg^{-1}(y)/dy$
  - Need to be able to calculate g^{-1}

Jensen’s inequality

- $E(g(X)) \neq g(E(X))$ unless g is linear
- If g convex up, E(g(X)) < g(E(X))



How about moments?

- Delta method approximation for E(g(X))
  - $g(x)=g(\mu)+g'(\mu)(x-\mu)+g''(\mu)(x-\mu)^2/2+\ldots$
  - $E(g(X)=E(g(\mu))+E(g'(\mu)(x-\mu))+E(g''(\mu)(x-\mu)^2/2)+\ldots$
  - $=g(\mu)+g'(\mu)E(X-\mu)+g''(\mu)/2E((x-\mu)^2)+\ldots$
  - $\approx g(m)+g''(m)Var(X)/2$
- No simple method for Var(g(X))

Delta example – Bet-hedging

- In a fluctuating environment: i.e. λ varies year-to-year : e.g. desert annuals (wet/dry years)
- $N_\tau = \lambda_1 \lambda_2 \ldots \lambda_t N_0$
- So $\lambda_{longterm} = (\lambda_1 \lambda_2 \ldots \lambda_t)^{1/t}$ or $\log \lambda_{longterm} = 1/t\ (\Sigma \log \lambda_i)$
- I.e. fitness = $E(\log \lambda_i)$
- Say we know distribution of λ_i (50% good/50% bad or lognormal or ...), what is average fitness?
  - Jensen's inequality: $\log(\lambda_{longterm}) < \log E(\lambda_i) = \log(\lambda_{average})$
  - Delta method approximates how much less: $g'(x)=1/x$ so:
    - $W \approx g(m)+g''(m)Var(X)/2 = \log \lambda_{avg} - Var(\lambda_{avg})/\lambda_{avg}^2$
- → Bet-hedging reduce variance even if a reduced mean:
  - 4/6 off spring better than 0/12 offspring in good/bad years

Limit

- Take n samples from a distribution
- As n gets large:
  - $Sum(X_i) \rightarrow N(n\mu, \sigma\sqrt{n})$
  - $Prod(x_i) \rightarrow$ Lognormal
  - $Max(x_i) \rightarrow$ Extreme value (log-Weibull)
    - Example coldest frost temperature
  - $Min(x_i) \rightarrow$ -Extreme value

Spurious correlations (Brett)

- Study of $\rho(X,X+Y)$, $\rho(X/Z,Y/Z)$, etc
- These are almost always somewhat correlated
- Traditional testing methods assume independence which fails here!
  - p values from traditional approaches useless
- Brett presents method – Monte Carlo
- Are these correlations spurious?
  - No often interesting biology
  - But if you report an r or p value you need to do the Monte Carlo

Repeated tests (Garcia)

- Say I measure 20 traits of a flower in two distinct populations
- I want to suggest that the populations are morphologically distinct
- What if:
  - 1 trait is significant at $p<0.05$?
  - 18 traits are significant at $p<0.05$?
  - 3 traits are significant at $p<0.05$?
- Net:
  - Bonferroni ($p<\alpha/n$) is overly conservative and overused
  - But do have to think about this issue

Repeated tests

k	p	Wrong psig	Bonf Psig	B&H psig	Result
6	0.006	0.05	0.05/6= 0.00833	0.05*6/6 =0.05	S
5	0.01	0.05	0.00833	0.04167	S
4	0.02	0.05	0.00833	0.3333	S
3	0.03	0.05	0.00833	0.025	NS
2	0.04	0.05	0.00833	0.1667	NS
1	0.06	0.05	0.00833	0.008333	NS

Summary

- Can sometimes avoid issue by just doing calculations on sampled data (item by item)
- But for any theoretical work/expectations need more sophisticated approach
  - Distribution of X*Y complicated (even (X+Y)
  - Moments are simple IF independent
  - Distribution of g(X) complicated
    - E(g(X)) delta approximation