Advanced Quantitative Research Methodology, Lecture Notes: Introduction[1]

Gary King
http://GKing.Harvard.edu

January 28, 2013

Who Takes This Course?

- Most Gov Dept grad students doing empirical work, the 2nd course in their methods sequence (Gov2001)
- Grad students from other departments (Gov2001)
- Undergrads (Gov1002)
- Non-Harvard students, visitors, faculty, & others (online through the Harvard Extension school, E-2001)
- Some of the best experiences here: getting to know people in other fields

How much math will you scare us with?

- All math requires two parts: proof and concepts & intuition
- Different classes emphasize:
  - rigorous abstract proofs (some)
  - dumbed down proofs, vague intuition (many)
  - us: deep concepts and intuition, proofs when needed (few)
- This class:
  - Overall goal: how to do empirical research, in depth
  - Use abstract statistical theory — when needed
  - Insure we understand the intuition — always
  - Always traverse from theoretical foundations to practical applications
  - $\rightsquigarrow$ Fewer proofs, more concepts, much more practical knowledge
- Do you have the background: What's this?

$$b = (X'X)^{-1}X'y$$

What's this Course About?

- Specific statistical methods for many research problems
  - How to learn (or create) new methods
  - Inference: Using facts you know to learn about facts you don't know
- How to write a publishable scholarly paper
- All the practical tools of research — theory, applications, simulation, programming, word processing, plumbing, whatever is useful
- Outline and class materials:

j.mp/G2001

  - The syllabus gives topics, not a weekly plan.
  - We will go as fast as possible subject to everyone following along
  - We cover different amounts of material each week

Requirements

1. Weekly assignments
   - Readings & videos with annotations, and assignments
   - Take notes, read carefully, don't skip equations
2. One "publishable" coauthored paper. (Easier than you think!)
   - Many class papers have been published, presented at conferences, become dissertations or senior theses, and won many awards
   - Undergrads have often had professional journal publications
   - Draft submission and replication exercise helps a lot.
   - See "Publication, Publication"
3. Participation and collaboration:
   - Do assignments in groups: "Cheating" is encouraged, so long as you write up your work on your own.
   - Participating in a conversation $>>$ Evesdropping
   - Use collaborative learning tools (we'll introduce)
   - Build class camaraderie: prepare, participate, help others
4. Focus, like I will, on what you learn, not your grades: Especially when we work on papers, I will treat you like a colleague, not a student

Got Questions?

- Send and respond to emails: gov2001-l@lists.fas.harvard.edu
- Browse archive of previous year's emails (Note which now-famous scholar is asking the question!)
- Q&A annotations in videos, readings, and assignments
- In-ter-rupt me as often as necessary
- Got a dumb question? Assume you are the smartest person in class and you eventually will be!
- When are Gary's office hours?

What is the field of statistics?

- The field of statistics originates in the study of politics and government: "*state*-istics", (circa 1662)
- A new field: Random assignment dates to the mid-1930s.
- The modern theory of inference dates only to the 1950s.
- Part of a monumental societal change, the march of quantification through academic, professional, commercial, and policy fields. (Popular books: *The Numerati*, *SuperCrunchers*, *MoneyBall*)
- The number of new methods is increasing fast
- Most important methods originate *outside* the discipline of statistics (random assignment, experimental design, survey research, machine learning, MCMC methods, . . .). Statistics: abstracts, proves formal properties, generalizes, and distributes results back out.

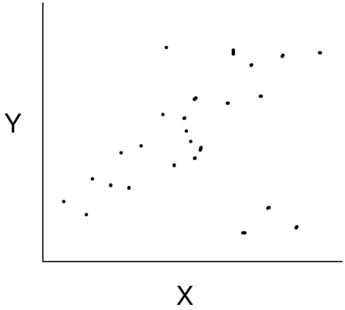
What is the subfield of political methodology?

- The methods subfield of political science, a relative of econometrics, psychological statistics, biostatistics, chemometrics, sociological methodology, cliometrics, stylometry, etc.
- Historically, political methodologists were trained in many different areas, and so the field is heavily interdisciplinary.
- As the cross-roads for other disciplines, it is one of the best places to learn about methods broadly. It reflects the diverse nature of political science.
- Second largest APSA Section, after the catchall Comparative Politics
- (Valuable for the job market!)
- Part of a massive change in the evidence base of the social sciences: (a) surveys, (b) end of period government stats, and (c) one-off studies of people, places, or events $\rightsquigarrow$ numerous new types and huge quantities of (big) data

Course strategy

- We could teach you the latest and greatest methods, but when you graduate they will be old
- We could teach you all the methods that might prove useful during your career, but when you graduate you will be old
- Instead, we teach you the fundamentals, the underlying theory of inference, from which most statistical models are developed.
- This helps us separate the conventions from underlying statistical theory. (How to get an F in Econometrics: follow advice from Psychometrics. Works in reverse too, even when the foundations are identical.)

e.g.,: How to fit a line to a scatterplot?

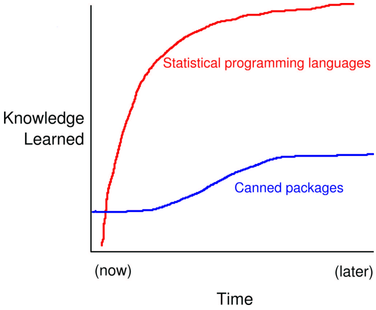


- visually (tends to be principle components)
- a rule (least squares, least absolute deviations, etc.)
- criteria (unbiasedness, efficiency, sufficiency, admissibility, etc.)
- from a theory of inference, and for a substantive purpose (like causal estimation, prediction, etc.)

The Fundamentals

- The fundamentals help us decide what is junk, new jargon, or a genuine advance
- We will reinvent existing methods by creating them from scratch.
- We will learn: its as easy to invent brand new methods too, when needed.
- The fundamentals help us pick up new methods easily.
- What's the "proper" way to teach statistics? Options:
  1. Years of calculus, real analysis, linear algebra, mathematical statistics, and probability theory. Then begin data analysis. (Works great, but not if you want to be a social scientist!)
  2. Teach the fundamentals, then do examples in great detail. Introduce math in almost as much depth, and only when needed.

Software options



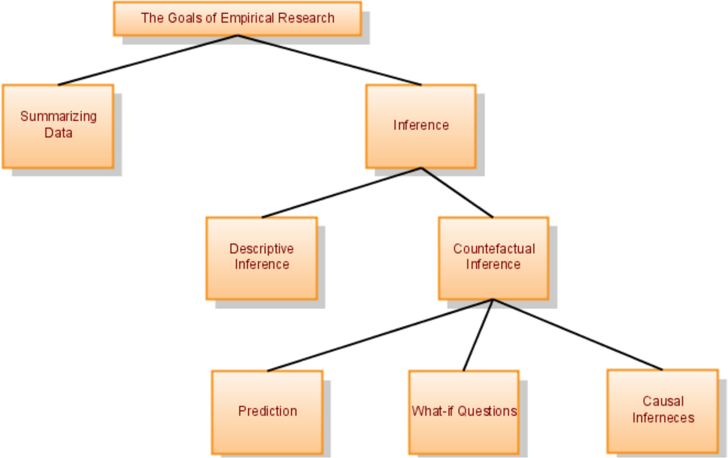
- We'll use R — a free open source program, a commons, a movement
- and an R program called Zelig (Imai, King, and Lau, 2006-12) which simplifies R and helps you up the steep slope fast (see j.mp/Zelig4)

What is this?

- Now you know what a model is. (Its an abstraction.)
- Is this model true?
- Are models ever true or false?
- Are models ever realistic or not?
- Are models ever useful or not?
- Models of dirt on airplanes, vs models of aerodynamics

Target Quantities of Interest

Inference (using facts you know to learn facts you don't know) v summarization



Statistical Models: Variable Definitions

- Explanatory variables (or "covariates," or "independent" or "exogenous" variables): $X = (x_1, x_2, \ldots, x_j, \ldots, x_k)$ for $x_j = \{x_{ij}\}$. X is $n \times k$.
- Dependent (or "outcome") variable: Y is $n \times 1$.
- Y_i, a random variable (before we know it)
- y_i, a number (after we know it)
- Common misunderstanding: a "dependent variable" can be
  - a column of numbers in your data set
  - the random variable *for each unit* i.
- X is fixed (not random).

Equivalent Linear Regression Notation

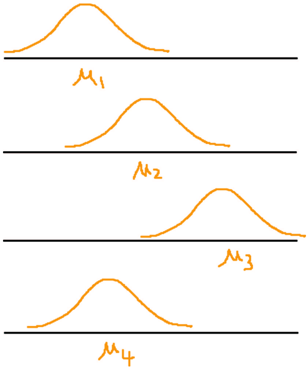
- Standard version:

$$\begin{aligned} Y_i &= x_i\beta + \epsilon_i && = \text{systematic} + \text{stochastic} \\ \epsilon_i &\sim f_N(e_i|0, \sigma^2) \end{aligned}$$

- Alternative version:

$$\begin{aligned} Y_i &\sim f_N(y_i|\mu_i, \sigma^2) && \text{stochastic} \\ \mu_i &= x_i\beta && \text{systematic} \end{aligned}$$

Understanding the Alternative Regression Notation



Is a histogram of y a test of normality?

Generalized Alternative Notation for Most Statistical Models

$$Y_i \sim f(y_i|\theta_i, \alpha) \qquad \text{stochastic}$$
$$\theta_i = g(X_i, \beta) \qquad \text{systematic}$$

where

Y_i random outcome variable

y_i realization of Y_i

$f(\cdot)$ probability density

θ_i a systematic feature of the density that varies over i

α ancillary parameter (feature of the density constant over i)

$g(\cdot)$ functional form

X_i explanatory variables

β effect parameters

Forms of Uncertainty

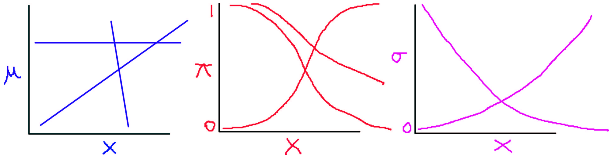
$$Y_i \sim f(y_i|\theta_i, \alpha) \qquad \text{stochastic}$$
$$\theta_i = g(X_i, \beta) \qquad \text{systematic}$$

- Estimation uncertainty: Lack of knowledge of β and α. Vanishes as n gets larger.
- Fundamental uncertainty: Represented by the stochastic component. Exists no matter what the researcher does; no matter how large n is.
- (If you know the model, is $R^2 = 1$? Can you predict Y perfectly?)

Systematic Components: Examples

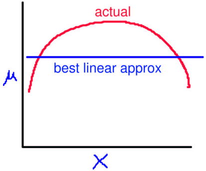
- $E(Y_i) \equiv \mu_i = X_i\beta = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}$
- $\Pr(Y_i = 1) \equiv \pi_i = \frac{1}{1+e^{-x_i\beta}}$
- $V(Y_i) \equiv \sigma_i^2 = e^{x_i\beta}$

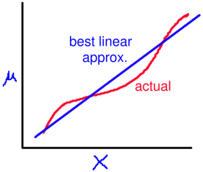
  (β is an "effect parameter" vector in each, but the meaning differs.)



- Each mathematical form is a class of functional forms
- We choose a member of the class by setting β

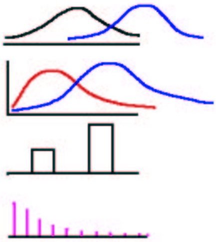
- We (ultimately) will
  - *Assume* (choose) one class of functional forms
  - Choose the member of the class by using data to *estimate* β
  - Since data contain (sampling, measurement, random) error, we will be uncertain to a degree about the member of the family (value of β).
- These forms are flexible and map many possible functional relationships
- If the true relationship falls outside the assumed class, we
  - Have specification error.
  - Get the best [linear,logit,etc] approximation to the correct functional form.
  - Depending on the case, this approximation may be close or far from the truth.





Overview of Stochastic Components: Describe the *sample space* (details shortly)

- Normal — continuous, unimodal, symmetric, unbounded
- Log-normal — continuous, unimodal, skewed, bounded from below by zero
- Bernoulli — discrete, binary outcomes
- Poisson — discrete, countably infinite on the nonnegative integers (for counts)



Choosing systematic and stochastic components

- If one is bounded, so is the other
- If the stochastic component is bounded, the systematic component must be (globally) nonlinear. (it could be locally linear)
- All modeling decisions can be decided if you know the data generation process — the whole process by which the data made its way from the world (including how the world produced the data) to your data set.
- What if we don't know the DGP (& we usually don't)?
  - The problem: model dependence
  - Our first approach: make "reasonable" assumptions and check fit (& other observable implications of the assumptions)
  - Later: relax functional form and distributional assumptions, or preprocess data (via matching, etc.) to avoid their consequences

Probability as a Model of Uncertainty

- $\Pr(y|M) = \Pr(\text{data}|\text{Model})$, where $M = (f, g, X, \beta, \alpha)$.
- 3 *axioms* define the function $\Pr(\cdot|\cdot)$:
  1. $\Pr(z) \geq 0$ for some event z
  2. $\Pr(\text{sample space}) = 1$
  3. If $z_1, \ldots, z_k$ are mutually exclusive events,

$$\Pr(z_1 \cup \cdots \cup z_k) = \Pr(z_1) + \cdots + \Pr(z_k),$$

- The first two imply $0 \leq \Pr(z) \leq 1$
- Axioms are not assumptions; they can't be wrong.
- From the axioms come *all* rules of probability theory.
- Rules can be applied *analytically* or via *simulation*.

Simulation is used to:

- solve probability problems
- evaluate estimators
- calculate features of probability densities
- transform statistical results into quantities of interest
- $\rightsquigarrow$ Experiments: students get the right answer far more frequently by using simulation than math

What is simulation?

Survey Sampling	Simulation
1. Learn about a population by taking a random sample from it	1. Learn about a distribution by taking random draws from it
2. Use the random sample to estimate a feature of the population	2. Use the random draws to approximate a feature of the distribution
3. The estimate is arbitrarily precise for large n	3. The approximation is arbitrarily precise for large M
4. Example: estimate the mean of the population	4. Example: Approximate the mean of the distribution

Simulation examples for solving probability problems

The Birthday Problem

Given a room with 24 randomly selected people, what is the probability that at least two have the same birthday?

```
sims <- 1000
people <- 24
alldays <- seq(1, 365, 1)
sameday <- 0
for (i in 1:sims) {
   room <- sample(alldays, people, replace = TRUE)
   if (length(unique(room)) < people) # same birthday
   sameday <- sameday+1
}

cat("Probability of >=2 people having the same birthday:", sameday/sims, "\n"
```

```
Four runs: .538, .550, .547, .524
```

Let's Make a Deal

In Let's Make a Deal, Monte Hall offers what is behind one of three doors. Behind a random door is a car; behind the other two are goats. You choose one door at random. Monte peeks behind the other two doors and opens the one (or one of the two) with the goat. He asks whether you'd like to switch your door with the other door that hasn't been opened yet. Should you switch?

```
    sims <- 1000
    WinNoSwitch <- 0
    WinSwitch <- 0
    doors <- c(1, 2, 3)
    for (i in 1:sims) {
      WinDoor <- sample(doors, 1)
      choice <- sample(doors, 1)
      if (WinDoor == choice)                       # no switch
      WinNoSwitch <- WinNoSwitch + 1
      doorsLeft <- doors[doors != choice]          # switch
      if (any(doorsLeft == WinDoor))
      WinSwitch <- WinSwitch + 1
    }
  cat("Prob(Car | no switch)=", WinNoSwitch/sims, "\n")
  cat("Prob(Car | switch)=", WinSwitch/sims, "\n")
```

Let's Make a Deal

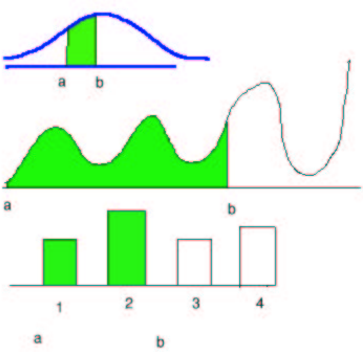
$\Pr(\text{car}\vert\text{No Switch})$	$\Pr(\text{car}\vert\text{Switch})$
.324	.676
.345	.655
.320	.680
.327	.673

What is a Probability Density?

A probability density is a function, $\mathrm{P}(Y)$, such that

1. Sum over all possible Y is 1.0
   - For discrete Y: $\sum_{\text{all possible} Y} \mathrm{P}(Y) = 1$
   - For continuous Y: $\int_{-\infty}^{\infty} \mathrm{P}(Y) dY = 1$
2. $\mathrm{P}(Y) \geq 0$ for every Y

Computing Probabilities from Densities



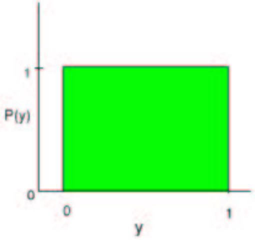
- For both: $\Pr(a \leq Y \leq b) = \int_a^b \mathrm{P}(Y)dY$
- For discrete: $\Pr(y) = \mathrm{P}(y)$
- For continuous: $\Pr(y) = 0$ (why?)

What you should know about every pdf

- The assignment of a probability or probability density to every conceivable value of Y_i
- The first principles
- How to use the final expression (but not necessarily the full derivation)
- How to simulate from the density
- How to compute features of the density such as its "moments"
- How to verify that the final expression is indeed a proper density

First Principles about the process that generates Y_i is such that

- Y_i falls in the interval $[0,1]$ with probability 1: $\int_0^1 \mathrm{P}(y)dy = 1$
- $\Pr(Y \in (a,b)) = \Pr(Y \in (c,d))$ if $a < b$, $c < d$, and $b - a = d - c$.

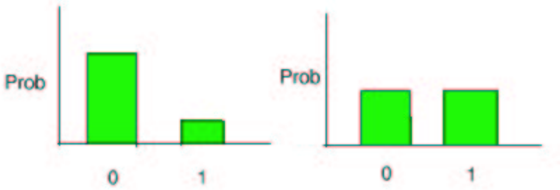


- Is it a pdf? How do you know?
- How to simulate? runif(1000)

Bernoulli pmf

- First principles about the process that generates Y_i:
  - Y_i has 2 mutually exclusive outcomes; and
  - The 2 outcomes are exhaustive
- In this simple case, we will compute features analytically and by simulation.
- Mathematical expression for the pmf
  - $\Pr(Y_i = 1|\pi_i) = \pi_i$, $\Pr(Y_i = 0|\pi_i) = 1 - \pi_i$
  - The parameter π happens to be interpretable as a probability
  - $\implies \Pr(Y_i = y|\pi_i) = \pi_i^y(1 - \pi_i)^{1-y}$
  - Alternative notation: $\Pr(Y_i = y|\pi_i) = \text{Bernoulli}(y|\pi_i) = f_b(y|\pi_i)$

Graphical summary of the Bernoulli



Expected value of the Bernoulli: analytically

- Expected value:

$$\begin{aligned} E(Y) &= \sum_{\text{all } y} y\text{P}(y) \\ &= 0\,\text{Pr}(0) + 1\,\text{Pr}(1) \\ &= \pi \end{aligned}$$

- Expected values of functions, $g(Y)$ of random variables Y

$$E[g(Y)] = \sum_{\text{all } y} g(y)\text{P}(y)$$

or

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)\text{P}(y)$$

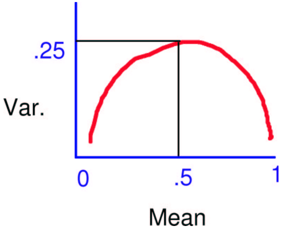
For example,

$$\begin{aligned} E(Y^2) &= \sum_{\text{all } y} y^2\text{P}(y) \\ &= 0^2\,\text{Pr}(0) + 1^2\,\text{Pr}(1) \\ &= \pi \end{aligned}$$

Variance of the Bernoulli (uses above results)

$$
\begin{aligned}
V(Y) &= E[(Y - E(Y))^2] && \text{(The definition)} \\
&= E(Y^2) - E(Y)^2 && \text{(An easier version)} \\
&= \pi - \pi^2 \\
&= \pi(1 - \pi)
\end{aligned}
$$

This makes sense:



How to Simulate from the Bernoulli with parameter π

- take one draw u from a uniform density on the interval [0,1]
- Set π to a particular value
- Set $y = 1$ if $u < \pi$ and $y = 0$ otherwise
- In R:

```
sims <- 1000     # set parameters
bernpi <- 0.2
u <- runif(sims)                # uniform sims
y <- as.integer(u < bernpi)
y                               # print results

Running the program gives:

0 0 0 1 0 0 1 1 0 0 1 1 1 0 ...
```

Binomial Distribution

First principles:

- N Bernoulli trials, $y_1, \dots, y_N$
- The trials are independent
- The trials are identically distributed
- We observe $Y = \sum_{i=1}^{N} y_i$

Density:

$$P(Y = y|\pi) = \binom{N}{y} \pi^y (1-\pi)^{N-y}$$

Explanation:

- $\binom{N}{y}$ because (1 0 1) and (1 1 0) are both $y = 2$.
- π^y because y successes with π probability each (product taken due to independence)
- $(1-\pi)^{N-y}$ because $N - y$ failures with $1 - \pi$ probability each
- Mean $E(Y) = N\pi$
- Variance $V(Y) = \pi(1-\pi)/N$.

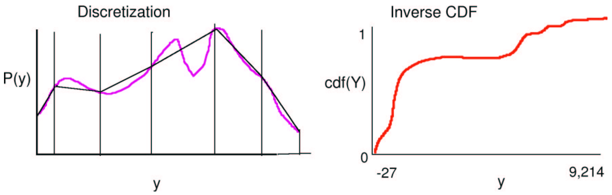
How to simulate from Binomial with parameter π and index N?

- Simulate N independent Bernoulli variables with parameter π
- Add them up

Where to get uniform random numbers

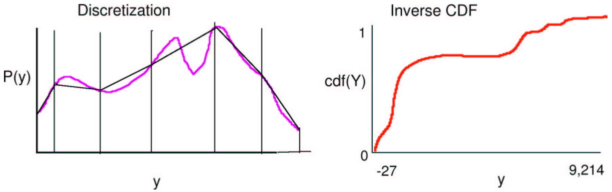
- Random is not haphazard (e.g., Benford's law)
- Random number generators are perfectly predictable (what?)
- We use pseudo-random numbers which have (a) digits that occur with 1/10th probability, (b) no time series patterns, etc.
- How to create real random numbers?
- Some chips now use quantum effects

"Discretization" for random draws from discrete pmfs, given uniform random numbers



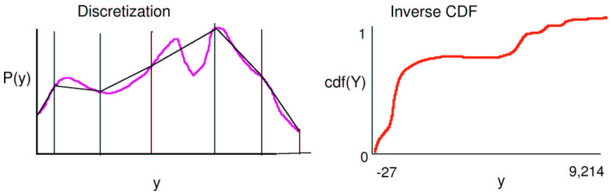
- Divide up PDF into a grid
- Approximate area (density/probability) above each interval
- Map [0,1] to the densities proportional to area
- Not feasible for too many dimensions

Inverse CDF method for random draws from continuous pdfs given uniform random numbers



- From the pdf $f(Y)$, compute the cdf: $\Pr(Y \leq y) \equiv F(y) = \int_{-\infty}^{y} f(z)dz$
- Define the inverse cdf $F^{-1}(y)$, such that $F^{-1}[F(y)] = y$
- Draw random uniform number, U
- Then $F^{-1}(U)$ gives a random draw from $f(Y)$.

A Refined Discretization method



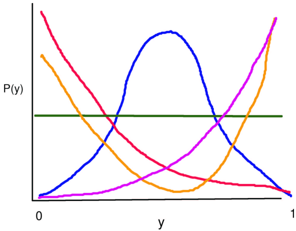
- Choose interval randomly as above
- Draw a number within an interval by the inverse CDF method applied to the trapezoidal approximation.
- Drawing random numbers from arbitrary multivariate densities: now an enormous literature

Stop here

We will stop here this year and skip to the next set of slides. Please refer to the notes below for further information on probability densities and random number generation.

Beta (continuous) density

- Used to model proportions.
- We'll use it first to generalize the Binomial distribution
- y falls in the interval [0,1]
- Takes on a variety of flexible forms, depending on the parameter values:



Standard Parameterization

$$\text{Beta}(y|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}$$

where, $\Gamma(x)$ is the gamma function:

$$\Gamma(x) = \int_0^\infty z^{x-1} e^{-z} dz$$

For integer values of x, $\Gamma(x+1) = x! = x(x-1)(x-2)\cdots 1$.

Non-integer values of x produce a continuous interpolation. In R or gauss: `gamma(x);`

Intuitive? The moments help some:

$$E(Y) = \frac{\alpha}{(\alpha+\beta)}$$

$$V(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Alternative parameterization

Set $\mu = E(Y) = \frac{\alpha}{(\alpha+\beta)}$ and $\frac{\mu(1-\mu)\gamma}{(1+\gamma)} = V(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$, solve for α and β and substitute in.

Result:

$$\text{beta}(y|\mu,\gamma) = \frac{\Gamma\left(\mu\gamma^{-1} + (1-\mu)\gamma^{-1}\right)}{\Gamma\left(\mu\gamma^{-1}\right)\Gamma\left[(1-\mu)\gamma^{-1}\right]} y^{\mu\gamma^{-1}-1}(1-y)^{(1-\mu)\gamma^{-1}-1}$$

where now $E(Y) = \mu$ and γ is an index of variation that varies with μ.

Reparameterization like this will be key throughout the course.

Beta-Binomial

Useful if the binomial variance is not approximately $\pi(1-\pi)/N$.

How to simulate

(First principles are easy to see from this too.)

- Begin with N Bernoulli trials with parameter π_j, $j = 1, \ldots, N$ (not necessarily independent or identically distributed)
- Choose $\mu = E(\pi_j)$ and γ
- Draw $\tilde{\pi}$ from Beta$(\pi|\mu, \gamma)$ (without this step we get Binomial draws)
- Draw N Bernoulli variables $\tilde{z}_j$ ($j = 1, \ldots, N$) from Bernoulli$(z_j|\tilde{\pi})$
- Add up the $\tilde{z}$'s to get $y = \sum_j^N \tilde{z}_j$, which is a draw from the beta-binomial.

Beta-Binomial Analytics

Recall:

$$\Pr(A|B) = \frac{\Pr(AB)}{\Pr(B)} \quad \Longrightarrow \quad \Pr(AB) = \Pr(A|B)\ Pr(B)$$

Plan:

- Derive the joint density of y and π. Then
- Average over the unknown π dimension

Hence, the beta-binomial (or extended beta-binomial):

$$\begin{aligned}
\mathrm{BB}(y_i|\mu,\gamma) &= \int_0^1 \mathrm{Binomial}(y_i|\pi) \times \mathrm{Beta}(\pi|\mu,\gamma) d\pi \\
&= \int_0^1 \mathrm{P}(y_i,\pi|\mu,\gamma) d\pi \\
&= \frac{N!}{y_i!(N-y_i)!} \prod_{j=0}^{y_i-1}(\mu+\gamma j) \prod_{j=0}^{N-y_i-1}(1-\mu+\gamma j) \prod_{j=0}^{N-1}(1+\gamma j)
\end{aligned}$$

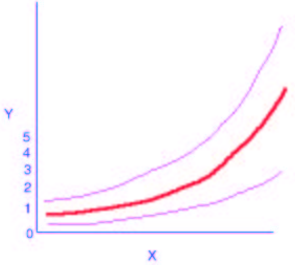
Poisson Distribution

- Begin with an observation period:

- All assumptions are about the events that occur between the start and when we observe the count. The process of event generation is assumed not observed.
- 0 events occur at the start of the period
- Only observe number of events at the end of the period
- No 2 events can occur at the same time
- Pr(event at time t | all events up to time $t-1$) is constant for all t.

Poisson Distribution

- First principles imply:
  - $$\text{Poisson}(y|\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} & \text{for } y_i = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$
  - $E(Y) = \lambda$
  - $V(Y) = \lambda$
  - That the variance goes up with the mean makes sense, but should they be equal?



Poisson Distribution

- If we assume Poisson dispersion, but $Y|X$ is *over-dispersed*, standard errors are too small.
- If we assume Poisson dispersion, but $Y|X$ is *under-dispersed*, standard errors are too large.

- How to simulate? We'll use canned random number generators.

Gamma Density

- Used to model durations and other nonnegative variables
- We'll use first to generalize the Poisson
- Parameters: $\phi > 0$ is the mean and $\sigma^2 > 1$ is an index of variability.
- Moments: mean $E(Y) = \phi > 0$ and variance $V(Y) = \phi(\sigma^2 - 1)$

$$\text{gamma}(y|\phi, \sigma^2) = \frac{y^{\phi(\sigma^2-1)^{-1}-1} e^{-y(\sigma^2-1)^{-1}}}{\Gamma[\phi(\sigma^2-1)^{-1}](\sigma^2-1)^{\phi(\sigma^2-1)^{-1}}}$$

Negative Binomial

- Same logic as the beta-binomial generalization of the binomial
- Parameters $\phi > 0$ and dispersion parameter $\sigma^2 > 1$
- Moments: mean $E(Y) = \phi > 0$ and variance $V(Y) = \sigma^2\phi$
- Allows *over-dispersion*: $V(Y) > E(Y)$.
- As $\sigma^2 \to 1$, NegBin$(y|\phi, \sigma^2) \to$ Poisson$(y|\phi)$ (i.e., small σ^2 makes the variation from the gamma vanish)

How to simulate (and first principles)

- Choose $E(Y) = \phi$ and σ^2
- Draw $\tilde{\lambda}$ from gamma$(\lambda|\phi, \sigma^2)$.
- Draw Y from Poisson$(y|\tilde{\lambda})$, which gives one draw from the negative binomial.

Negative Binomial Derivation

Recall:

$$\Pr(A|B) = \frac{\Pr(AB)}{\Pr(B)} \implies \Pr(AB) = \Pr(A|B)\Pr(B)$$

$$\begin{aligned}
\text{NegBin}(y|\phi,\sigma^2) &= \int_0^\infty \text{Poisson}(y|\lambda) \times \text{gamma}(\lambda|\phi,\sigma^2) d\lambda \\
&= \int_0^\infty \text{P}(y,\lambda|\phi,\sigma^2) d\lambda \\
&= \frac{\Gamma\left(\frac{\phi}{\sigma^2-1} + y_i\right)}{y_i!\Gamma\left(\frac{\phi}{\sigma^2-1}\right)} \left(\frac{\sigma^2-1}{\sigma^2}\right)^{y_i} \left(\sigma^2\right)^{\frac{-\phi}{\sigma^2-1}}
\end{aligned}$$

Normal Distribution

- Many different first principles
- A common one is the central limit theorem
- The univariate normal density:

$$N(y_i|\mu_i, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left(\frac{-(y_i - \mu_i)^2}{2\sigma^2}\right)$$

- The stylized normal: $f_{stn}(y_i|\mu_i) = N(y|\mu_i, 1)$

$$f_{stn}(y|\mu_i) = (2\pi)^{-1/2} \exp\left(\frac{-(y_i - \mu_i)^2}{2}\right)$$

- The standardized normal: $f_{sn}(y_i) = N(y_i|0, 1) = \phi(y_i)$

$$f_{sn}(y_i) = (2\pi)^{-1/2} \exp\left(\frac{-y_i^2}{2}\right)$$

Multivariate Normal Distribution

- Let $Y_i \equiv \{Y_{1i}, \ldots, Y_{ki}\}$ be a $k \times 1$ vector, jointly random:

$$Y_i \sim N(y_i|\mu_i, \Sigma)$$

where μ_i is $k \times 1$ and Σ is $k \times k$. For $k = 2$,

$$\mu_i = \begin{pmatrix} \mu_{1i} \\ \mu_{2i} \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

- Mathematical form:

$$N(y_i|\mu_i, \Sigma) = (2\pi)^{-k/2}|\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(y_i - \mu_i)'\Sigma^{-1}(y_i - \mu_i)\right]$$

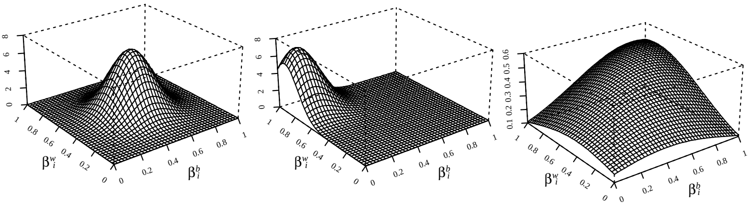
- Simulating *once* from this density produces k numbers. Special algorithms are used to generate normal random variates (in R, `mvrnorm()`, from the `MASS` library).

Multivariate Normal Distribution

- Moments: $E(Y) = \mu_i$, $V(Y) = \Sigma$, $\text{Cov}(Y_1, Y_2) = \sigma_{12} = \sigma_{21}$.
- $\text{Corr}(Y_1, Y_2) = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$
- Marginals:

$$N(Y_1|\mu_1, \sigma_1^2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} N(y_i|\mu_i, \Sigma) dy_2 dy_3 \cdots dy_k$$

Truncated bivariate normal examples (for β^b and β^w)



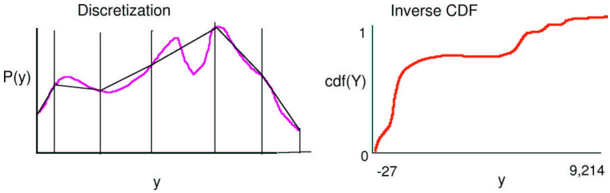
(a) 0.5 0.5 0.15 0.15 0 (b) 0.1 0.9 0.15 0.15 0 (c) 0.8 0.8 0.6 0.6 0.5

Parameters are μ_1, μ_2, σ_1, σ_2, and ρ.

Where to get uniform random numbers

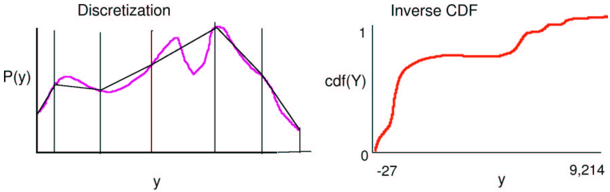
- Random is not haphazard (e.g., Benford's law)
- Computer random number generators are perfectly predictable.
- We use pseudo-random numbers which have (a) digits that occur with 1/10th probability, (b) no time series patterns, etc.
- How to create real random numbers?
- Some chips now use quantum effects to create real random numbers.

"Discretization" for random draws from discrete pmfs, given uniform random numbers



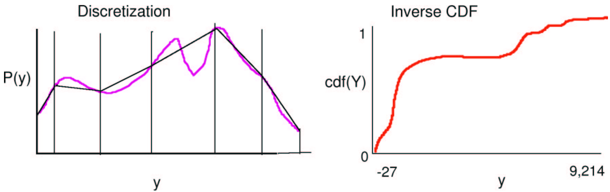
- Divide up PDF into a grid
- Compute by linear approximation area (density/probability) in each interval
- Map [0,1] to the densities proprtionally
- Not feasible for multivariate random number generation

Inverse CDF method for random draws from continuous pdfs given uniform random numbers



- From the pdf $f(Y)$, compute the cdf: $\Pr(Y \leq y) \equiv F(y) = \int_{-\infty}^{y} f(z)dz$
- Define the inverse cdf $F^{-1}(y)$, such that $F^{-1}[F(y)] = y$
- Draw random uniform number, U
- Then $F^{-1}(U)$ gives a random draw from $f(Y)$.

A Refined Discretization method



- Choose interval randomly as above
- Draw a number within an interval by the inverse CDF method applied to the trapezoidal approximation.