Soc504: Basics

Brandon Stewart[1]

Princeton

February 8, 2017

[1] These slides are a very lightly remixed version of slides for Gary King's course Gov2001 at Harvard.

Followup

Followup

- Questions?

Followup

- Questions?
- Replication Stories?

Simulation is used to:

Simulation is used to:

1. Solve probability problems

Simulation is used to:

1. Solve probability problems
2. Evaluate estimators

Simulation is used to:

1. Solve probability problems
2. Evaluate estimators
3. Calculate features of probability densities

Simulation is used to:

1. Solve probability problems
2. Evaluate estimators
3. Calculate features of probability densities
4. Transform statistical results into quantities of interest

Simulation is used to:

1. Solve probability problems
2. Evaluate estimators
3. Calculate features of probability densities
4. Transform statistical results into quantities of interest
5. Get the right answer: students get the right answer far more frequently by using simulation than math

What is simulation?

What is simulation?

Survey Sampling	Simulation

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

What is simulation?

Survey Sampling	Simulation
1. Learn about a population by taking a random sample from it 2. Use the random sample to estimate a feature of the population	1. Learn about a distribution by taking random draws from it 2. Use the random draws to approximate a feature of the distribution

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

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

Monty Hall's Let's Make a Deal

Monty Hall's Let's Make a Deal

You have a choice among 3 doors. Behind a random door is a car; behind the other two are goats. You choose one at random. Monty peeks behind the other two doors and opens the one (or one of the two) with the goat and asks if you'd like to switch your door for the other door that hasn't been opened yet. Should you switch?

Monty Hall's Let's Make a Deal

You have a choice among 3 doors. Behind a random door is a car; behind the other two are goats. You choose one at random. Monty peeks behind the other two doors and opens the one (or one of the two) with the goat and asks if you'd like to switch your door for the other door that hasn't been opened yet. Should you switch?

```
sims <- 1000
WinNoSwitch <- 0
WinSwitch <- 0
doors <- c(1, 2, 3)
```

Monty Hall's Let's Make a Deal

You have a choice among 3 doors. Behind a random door is a car; behind the other two are goats. You choose one at random. Monty peeks behind the other two doors and opens the one (or one of the two) with the goat and asks if you'd like to switch your door for 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
}
```

Monty Hall's Let's Make a Deal

You have a choice among 3 doors. Behind a random door is a car; behind the other two are goats. You choose one at random. Monty peeks behind the other two doors and opens the one (or one of the two) with the goat and asks if you'd like to switch your door for the other door that hasn't been opened yet. Should you switch?

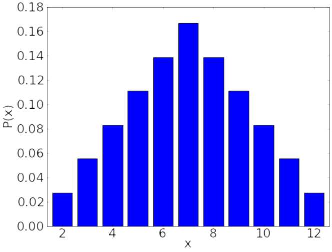
```
  sims <- 1000
  WinNoSwitch <- 0
  WinSwitch <- 0
  doors <- c(1, 2, 3)
  for (i in 1:sims) {
    WinDoor <- sample(doors, 1)
    choice <- sample(doors, 1)
    if (WinDoor == choice)                     # no switch
       WinNoSwitch <- WinNoSwitch + 1
    doorsLeft <- doors[doors != choice]  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

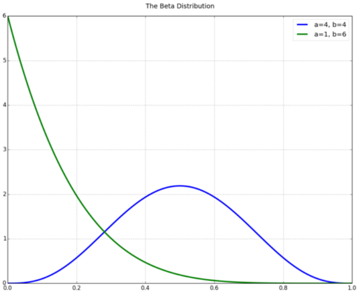
Let's Make a Deal

Pr(car|No Switch)	Pr(car|Switch)
.324	.676
.345	.655
.320	.680
.327	.673

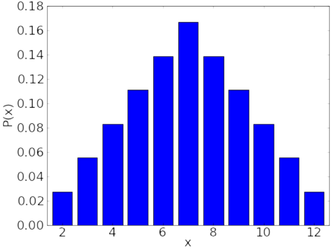
Computing Probabilities from PDFs

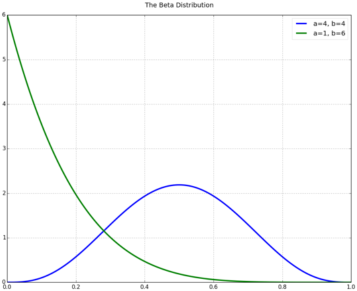
Computing Probabilities from PDFs





Computing Probabilities from PDFs

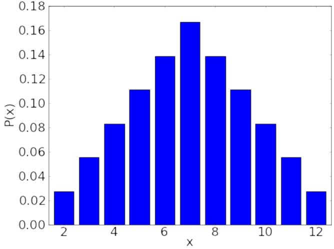


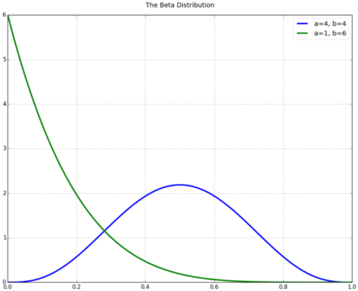


Required for a PDF:

Denoted: $\mathbb{P}(y)$

Computing Probabilities from PDFs



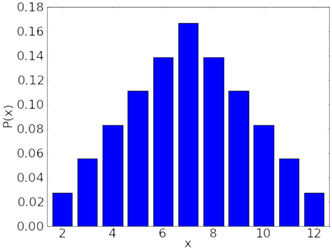


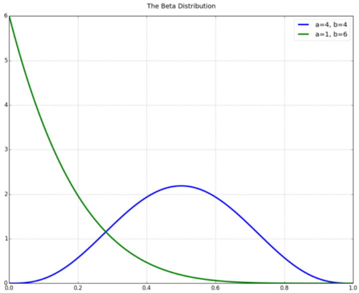
Required for a PDF:

Denoted: $\mathbb{P}(y)$

1. $\mathbb{P}(y) \geq 0$ for every y

Computing Probabilities from PDFs



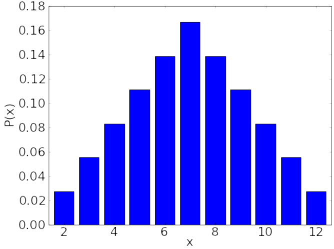


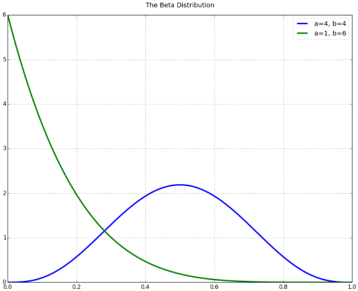
Required for a PDF:

Denoted: $\mathbb{P}(y)$

1. $\mathbb{P}(y) \geq 0$ for every y
2. $\sum_y \mathbb{P}(y) = 1$

Computing Probabilities from PDFs



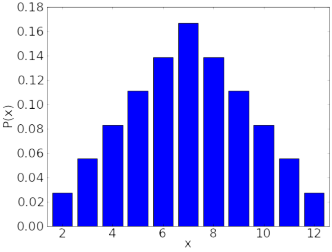


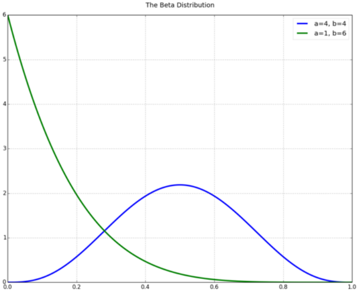
Required for a PDF:

Denoted: $\mathbb{P}(y)$

1. $\mathbb{P}(y) \geq 0$ for every y
2. $\sum_y \mathbb{P}(y) = 1$ or $\int_{-\infty}^{\infty} \mathbb{P}(y)dy = 1$

Computing Probabilities from PDFs





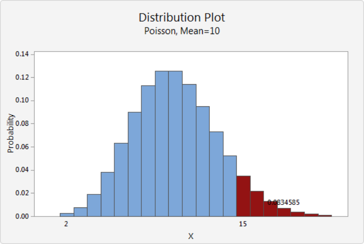
Required for a PDF:

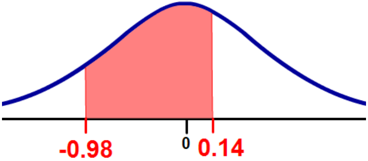
Denoted: $\mathbb{P}(y)$

1. $\mathbb{P}(y) \geq 0$ for every y
2. $\sum_y \mathbb{P}(y) = 1$ or $\int_{-\infty}^{\infty} \mathbb{P}(y)dy = 1$

Probability Calculations:

Computing Probabilities from PDFs





Required for a PDF:

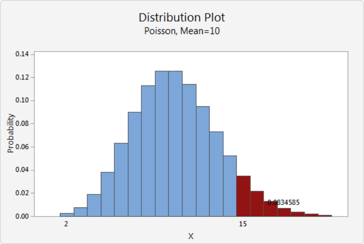
Denoted: $\mathbb{P}(y)$

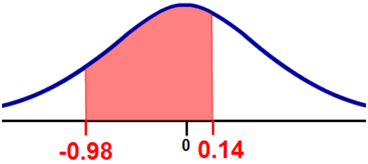
1. $\mathbb{P}(y) \geq 0$ for every y
2. $\sum_y \mathbb{P}(y) = 1$ or $\int_{-\infty}^{\infty} \mathbb{P}(y)dy = 1$

Probability Calculations:

- $\Pr(a \leq Y \leq b) = \int_a^b \mathbb{P}(y)dy$

Computing Probabilities from PDFs





Required for a PDF:

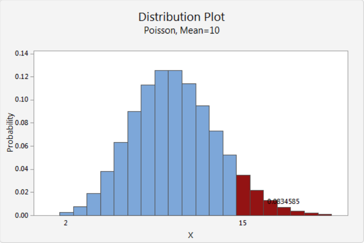
Denoted: $\mathbb{P}(y)$

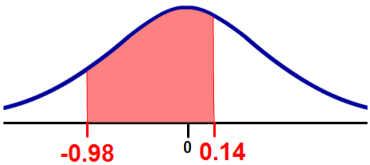
1. $\mathbb{P}(y) \geq 0$ for every y
2. $\sum_y \mathbb{P}(y) = 1$ or $\int_{-\infty}^{\infty} \mathbb{P}(y)dy = 1$

Probability Calculations:

- $\Pr(a \leq Y \leq b) = \int_a^b \mathbb{P}(y)dy$
- For discrete: $\Pr(y) = \mathbb{P}(y)$

Computing Probabilities from PDFs





Required for a PDF:

Denoted: $\mathbb{P}(y)$

1. $\mathbb{P}(y) \geq 0$ for every y
2. $\sum_y \mathbb{P}(y) = 1$ or $\int_{-\infty}^{\infty} \mathbb{P}(y)dy = 1$

Probability Calculations:

- $Pr(a \leq Y \leq b) = \int_a^b \mathbb{P}(y)dy$
- For discrete: $Pr(y) = \mathbb{P}(y)$
- For continuous: $Pr(y) = 0$

What you should know about every pdf

What you should know about every pdf

- The assignment of a probability or probability density to every conceivable value of Y_i

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

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)

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

What you should know about every pdf

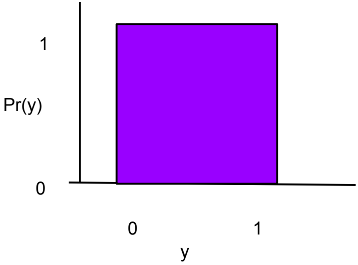
- The assignment of a probability or probability density to every conceivable value of Y_i
- The first principles
- How to use the final expression (but not necessarily the full derivation)
- How to simulate from the density
- How to compute features of the density such as its "moments"

What you should know about every pdf

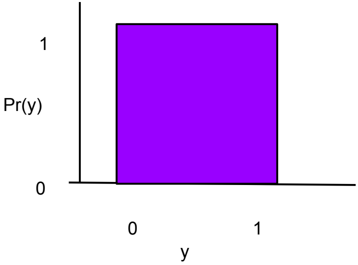
- The assignment of a probability or probability density to every conceivable value of Y_i
- The first principles
- How to use the final expression (but not necessarily the full derivation)
- How to simulate from the density
- How to compute features of the density such as its "moments"
- How to verify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

Uniform Density on the interval $[0, 1]$

Uniform Density on the interval $[0, 1]$

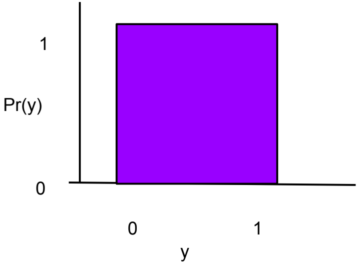


Uniform Density on the interval $[0, 1]$



First Principles about the process that generates Y_i is such that

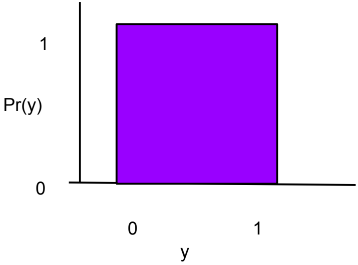
Uniform Density on the interval $[0, 1]$



First Principles about the process that generates Y_i is such that

- Y_i falls in the interval $[0, 1]$ with probability 1: $\int_0^1 \mathbb{P}(y)dy = 1$

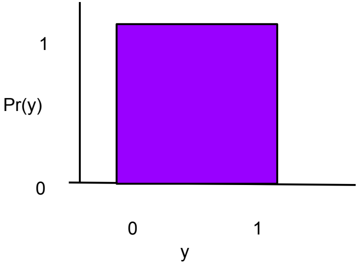
Uniform Density on the interval $[0, 1]$



First Principles about the process that generates Y_i is such that

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

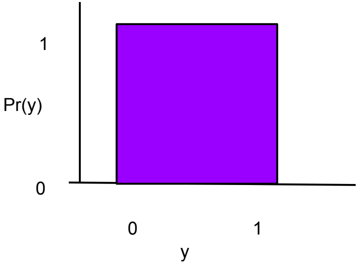
Uniform Density on the interval $[0, 1]$



First Principles about the process that generates Y_i is such that

- Y_i falls in the interval $[0, 1]$ with probability 1: $\int_0^1 \mathbb{P}(y)dy = 1$
- $\Pr(Y \in (a, b)) = \Pr(Y \in (c, d))$ if $a < b$, $c < d$, and $b - a = d - c$.
- Is it a pdf? How do you know?

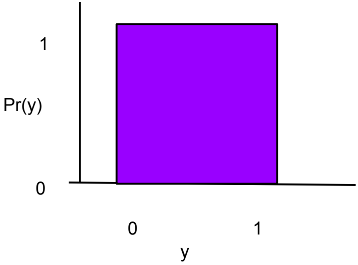
Uniform Density on the interval $[0, 1]$



First Principles about the process that generates Y_i is such that

- Y_i falls in the interval $[0, 1]$ with probability 1: $\int_0^1 \mathbb{P}(y)dy = 1$
- $\Pr(Y \in (a, b)) = \Pr(Y \in (c, d))$ if $a < b$, $c < d$, and $b - a = d - c$.
- Is it a pdf? How do you know?
- How to simulate?

Uniform Density on the interval $[0, 1]$

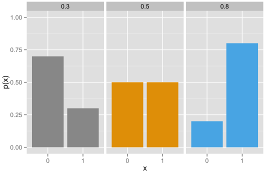


First Principles about the process that generates Y_i is such that

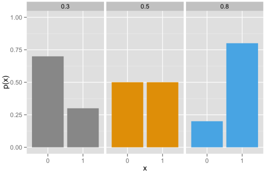
- Y_i falls in the interval $[0, 1]$ with probability 1: $\int_0^1 \mathbb{P}(y)dy = 1$
- $\Pr(Y \in (a, b)) = \Pr(Y \in (c, d))$ if $a < b$, $c < d$, and $b - a = d - c$.
- Is it a pdf? How do you know?
- How to simulate? runif(1000)

Bernoulli pmf

Bernoulli pmf

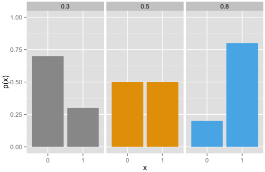


Bernoulli pmf



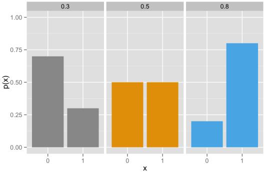
- First principles about the process that generates Y_i:

Bernoulli pmf



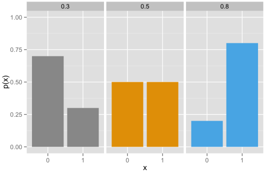
- First principles about the process that generates Y_i:
  - Y_i has 2 mutually exclusive outcomes; and

Bernoulli pmf



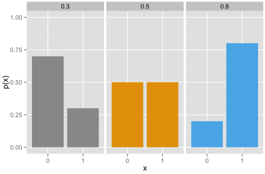
- First principles about the process that generates Y_i:
  - Y_i has 2 mutually exclusive outcomes; and
  - The 2 outcomes are exhaustive

Bernoulli pmf



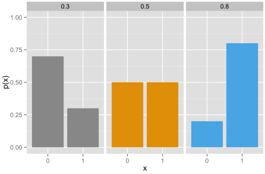
- First principles about the process that generates Y_i:
  - Y_i has 2 mutually exclusive outcomes; and
  - The 2 outcomes are exhaustive
- In this simple case, we will compute features analytically and by simulation.

Bernoulli pmf



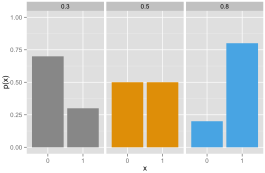
- First principles about the process that generates Y_i:
  - Y_i has 2 mutually exclusive outcomes; and
  - The 2 outcomes are exhaustive
- In this simple case, we will compute features analytically and by simulation.
- Mathematical expression for the pmf

Bernoulli pmf



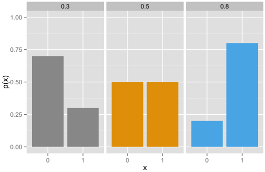
- First principles about the process that generates Y_i:
  - Y_i has 2 mutually exclusive outcomes; and
  - The 2 outcomes are exhaustive
- In this simple case, we will compute features analytically and by simulation.
- Mathematical expression for the pmf
  - $\Pr(Y_i = 1|\pi_i) = \pi_i$, $\Pr(Y_i = 0|\pi_i) = 1 - \pi_i$

Bernoulli pmf



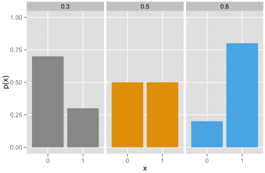
- First principles about the process that generates Y_i:
  - Y_i has 2 mutually exclusive outcomes; and
  - The 2 outcomes are exhaustive
- In this simple case, we will compute features analytically and by simulation.
- Mathematical expression for the pmf
  - $\Pr(Y_i = 1|\pi_i) = \pi_i$, $\Pr(Y_i =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

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
  - $\implies \Pr(Y_i = y|\pi_i) = \pi_i^y(1-\pi_i)^{1-y}$

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: $\Pr(Y_i = y|\pi_i) = \text{Bernoulli}(y|\pi_i) = f_b(y|\pi_i)$

Features of the Bernoulli: analytically

Features of the Bernoulli: analytically

- Expected value:

Features of the Bernoulli: analytically

- Expected value:

$$E(Y) = \sum_{\text{all } y} y\mathbb{P}(y)$$

Features of the Bernoulli: analytically

- Expected value:

$$E(Y) = \sum_{\text{all } y} y\mathbb{P}(y)$$
$$= 0\,\Pr(0) + 1\,\Pr(1)$$

Features of the Bernoulli: analytically

- Expected value:

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

Features of the Bernoulli: analytically

- Expected value:

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

- Variance:

Features of the Bernoulli: analytically

- Expected value:

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

- Variance:

$$V(Y) = E[(Y - E(Y))^2] \qquad \text{(The definition)}$$

Features of the Bernoulli: analytically

- Expected value:

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

- Variance:

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

Features of the Bernoulli: analytically

- Expected value:

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

- Variance:

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

Features of the Bernoulli: analytically

- Expected value:

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

- Variance:

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

- How do we compute $E(Y^2)$?

Expected values of functions of random variables

Expected values of functions of random variables

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

Expected values of functions of random variables

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

or

Expected values of functions of random variables

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

or

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

Expected values of functions of random variables

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

or

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

For example,

Expected values of functions of random variables

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

or

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

For example,

$$E(Y^2) = \sum_{\text{all } y} y^2\mathbb{P}(y)$$

Expected values of functions of random variables

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

or

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

For example,

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

Expected values of functions of random variables

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

or

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

For example,

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

Variance of the Bernoulli (uses above results)

Variance of the Bernoulli (uses above results)

$$V(Y) = E[(Y - E(Y))^2] \qquad \text{(The definition)}$$

Variance of the Bernoulli (uses above results)

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

Variance of the Bernoulli (uses above results)

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

Variance of the Bernoulli (uses above results)

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

Variance of the Bernoulli (uses above results)

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

Think about where the maximum is. Does it accord with your intuition?

How to Simulate from the Bernoulli with parameter π

How to Simulate from the Bernoulli with parameter π

- take one draw u from a uniform density on the interval [0,1]

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

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

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:

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
```

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:
```

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 ...
```

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 ...
```

- In practice, can use `rbinom(size=1)`

Binomial Distribution

Binomial Distribution

Binomial Distribution

First principles:

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

Binomial Distribution

First principles:

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

Density:

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

Binomial Distribution

First principles:

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

Density:

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

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

Binomial Distribution

First principles:

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

Density:

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

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 the Binomial distribution

How to simulate from the Binomial distribution

- To simulate from the Binomial(π; N):

How to simulate from the Binomial distribution

- To simulate from the Binomial(π; N):
  - Simulate N independent Bernoulli variables, $Y_1, \ldots, Y_N$, each with parameter π

How to simulate from the Binomial distribution

- To simulate from the Binomial(π; N):
  - Simulate N independent Bernoulli variables, $Y_1, \ldots, Y_N$, each with parameter π
  - Add them up: $\sum_{i=1}^{N} Y_i$

How to simulate from the Binomial distribution

- To simulate from the Binomial(π; N):
  - Simulate N independent Bernoulli variables, $Y_1, \ldots, Y_N$, each with parameter π
  - Add them up: $\sum_{i=1}^{N} Y_i$
  - `rbinom()` (which can also be used for Bernoulli)

How to simulate from the Binomial distribution

- To simulate from the Binomial(π; N):
  - Simulate N independent Bernoulli variables, $Y_1, \ldots, Y_N$, each with parameter π
  - Add them up: $\sum_{i=1}^{N} Y_i$
  - `rbinom()` (which can also be used for Bernoulli)
- What can you do with the simulations?

Where to get uniform random numbers

Where to get uniform random numbers

- Random is not haphazard (e.g., Benford's law)

Where to get uniform random numbers

- Random is not haphazard (e.g., Benford's law)
- Random number generators are perfectly predictable (what?)

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 $1/10$th probability, (b) no time series patterns, etc.

Where to get uniform random numbers

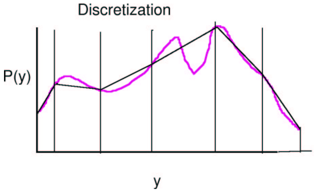
- Random is not haphazard (e.g., Benford's law)
- Random number generators are perfectly predictable (what?)
- We use pseudo-random numbers which have (a) digits that occur with 1/10th probability, (b) no time series patterns, etc.
- How to create real random numbers?

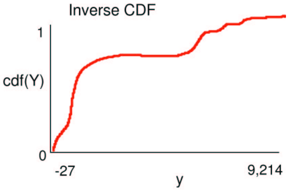
Where to get uniform random numbers

- 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?
- In the midrange future we might be using quantum computers for this.

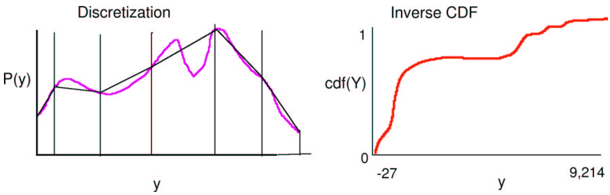
Discretization for random draws from discrete pmfs

Discretization for random draws from discrete pmfs



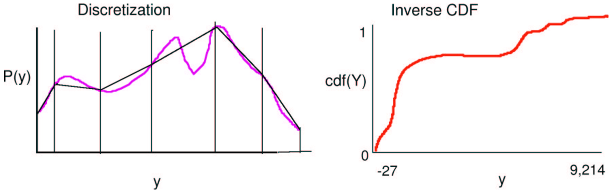


Discretization for random draws from discrete pmfs



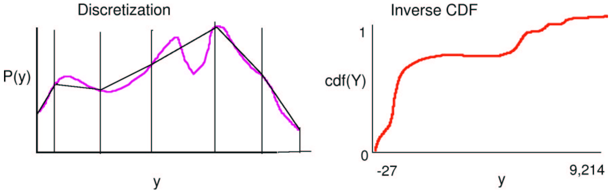
- Divide up PDF into a grid

Discretization for random draws from discrete pmfs



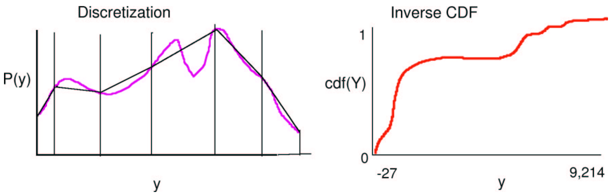
- Divide up PDF into a grid
- Approximate probabilities by trapezoids

Discretization for random draws from discrete pmfs



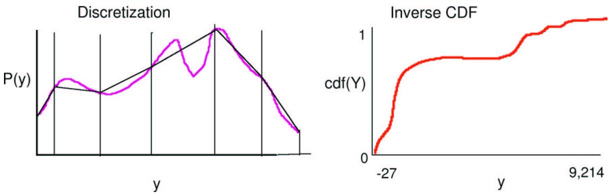
- Divide up PDF into a grid
- Approximate probabilities by trapezoids
- Map [0,1] uniform draws to the proportion area in each trapezoid

Discretization for random draws from discrete pmfs



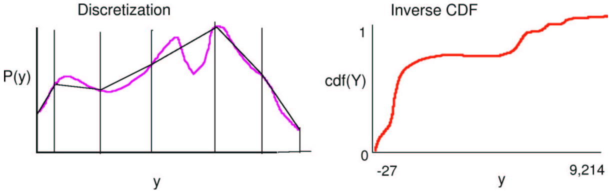
- Divide up PDF into a grid
- Approximate probabilities by trapezoids
- Map [0,1] uniform draws to the proportion area in each trapezoid
- Return midpoint of each trapezoid

Discretization for random draws from discrete pmfs



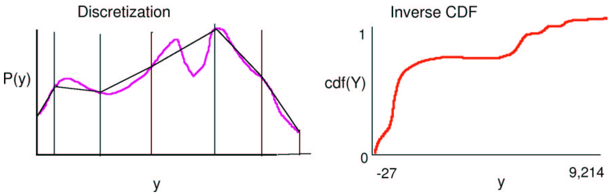
- Divide up PDF into a grid
- Approximate probabilities by trapezoids
- Map [0,1] uniform draws to the proportion area in each trapezoid
- Return midpoint of each trapezoid
- More trapezoids $\rightsquigarrow$ better approximation

Discretization for random draws from discrete pmfs

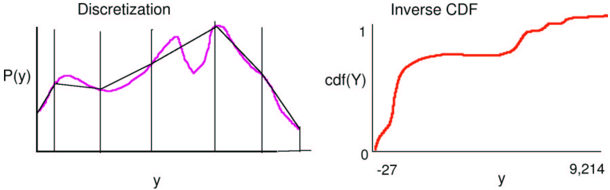


- Divide up PDF into a grid
- Approximate probabilities by trapezoids
- Map [0,1] uniform draws to the proportion area in each trapezoid
- Return midpoint of each trapezoid
- More trapezoids $\rightsquigarrow$ better approximation
- (Works for a few dimensions, but infeasible for many)

Inverse CDF: drawing from arbitrary continuous pdfs

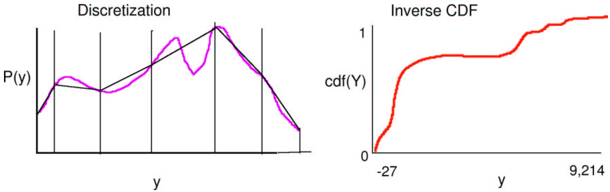


Inverse CDF: drawing from arbitrary continuous pdfs



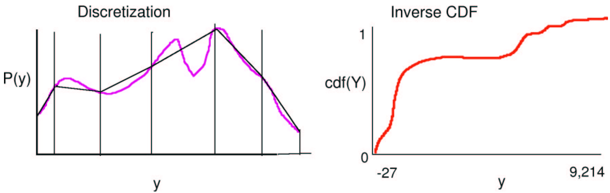
- From the pdf $f(Y)$, compute the cdf: $\Pr(Y \leq y) \equiv F(y) = \int_{-\infty}^{y} f(z)dz$

Inverse CDF: drawing from arbitrary continuous pdfs



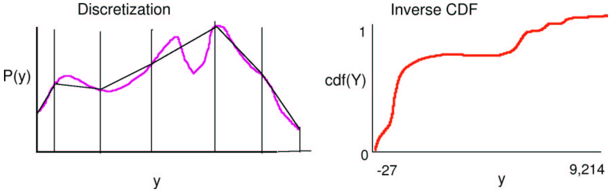
- From the pdf $f(Y)$, compute the cdf: $\Pr(Y \leq y) \equiv F(y) = \int_{-\infty}^{y} f(z)dz$
- Define the inverse cdf $F^{-1}(y)$, such that $F^{-1}[F(y)] = y$

Inverse CDF: drawing from arbitrary continuous pdfs



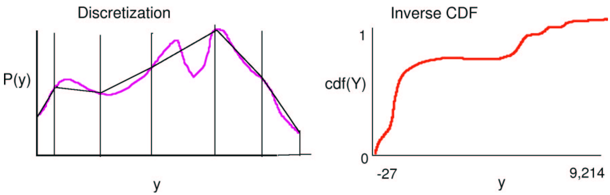
- From the pdf $f(Y)$, compute the cdf: $\Pr(Y \leq y) \equiv F(y) = \int_{-\infty}^{y} f(z)dz$
- Define the inverse cdf $F^{-1}(y)$, such that $F^{-1}[F(y)] = y$
- Draw random uniform number, U

Inverse CDF: drawing from arbitrary continuous pdfs

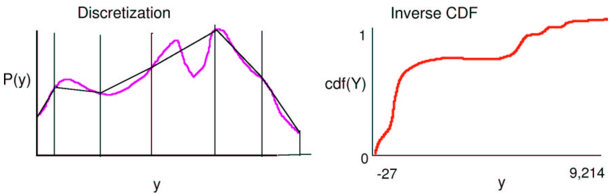


- From the pdf $f(Y)$, compute the cdf: $\Pr(Y \leq y) \equiv F(y) = \int_{-\infty}^{y} f(z)dz$
- Define the inverse cdf $F^{-1}(y)$, such that $F^{-1}[F(y)] = y$
- Draw random uniform number, U
- Then $F^{-1}(U)$ gives a random draw from $f(Y)$.

Using Inverse CDF to Improve Discretization Method

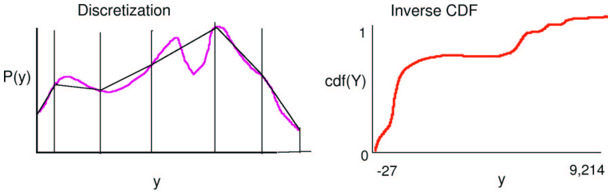


Using Inverse CDF to Improve Discretization Method



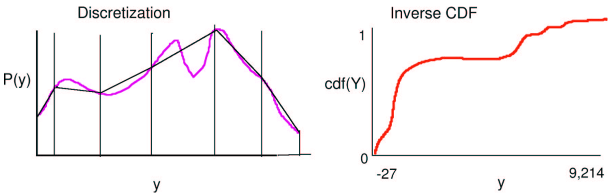
- Refined Discretization Method:

Using Inverse CDF to Improve Discretization Method



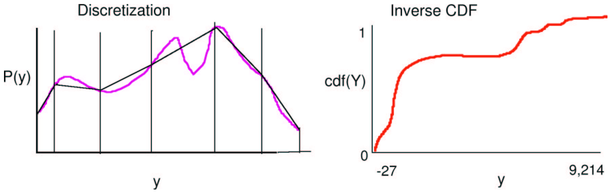
- Refined Discretization Method:
  - Choose interval randomly as above (based on area in trapezoids)

Using Inverse CDF to Improve Discretization Method



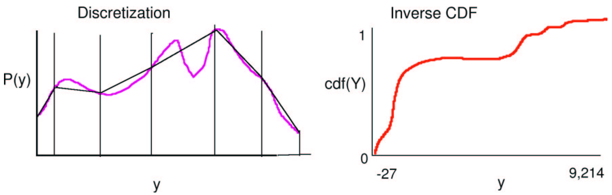
- Refined Discretization Method:
  - Choose interval randomly as above (based on area in trapezoids)
  - Draw a number within each trapezoid by the inverse CDF method applied to the trapezoidal approximation.

Using Inverse CDF to Improve Discretization Method



- Refined Discretization Method:
  - Choose interval randomly as above (based on area in trapezoids)
  - Draw a number within each trapezoid by the inverse CDF method applied to the trapezoidal approximation.
- Drawing random numbers from arbitrary multivariate densities: now an enormous literature

Using Inverse CDF to Improve Discretization Method



- Refined Discretization Method:
  - Choose interval randomly as above (based on area in trapezoids)
  - Draw a number within each trapezoid by the inverse CDF method applied to the trapezoidal approximation.
- Drawing random numbers from arbitrary multivariate densities: now an enormous literature
- Also a decent literature on drawing samples with different speed/accuracy tradeoffs

Normal Distribution

Normal Distribution

- Many different first principles

Normal Distribution

- Many different first principles
- A common one is the central limit theorem

Normal Distribution

- Many different first principles
- A common one is the central limit theorem
- The univariate normal density (with mean μ_i, variance σ^2)

Normal Distribution

- Many different first principles
- A common one is the central limit theorem
- The univariate normal density (with mean μ_i, variance σ^2)

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

Normal Distribution

- Many different first principles
- A common one is the central limit theorem
- The univariate normal density (with mean μ_i, variance σ^2)

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

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

Normal Distribution

- Many different first principles
- A common one is the central limit theorem
- The univariate normal density (with mean μ_i, variance σ^2)

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

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

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

Normal Distribution

- Many different first principles
- A common one is the central limit theorem
- The univariate normal density (with mean μ_i, variance σ^2)

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

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

Normal Distribution

- Many different first principles
- A common one is the central limit theorem
- The univariate normal density (with mean μ_i, variance σ^2)

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

- 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

Multivariate Normal Distribution

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

Multivariate Normal Distribution

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

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

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 $k = 2$,

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 $k = 2$,

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

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:

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:

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

Multivariate Normal Distribution

- Let $Y_i \equiv \{Y_{1i}, \dots, Y_{ki}\}$ 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 <u>once</u> 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: $E(Y) = \mu_i$, $V(Y) = \Sigma$, $\text{Cov}(Y_1, Y_2) = \sigma_{12} = \sigma_{21}$.

Multivariate Normal Distribution

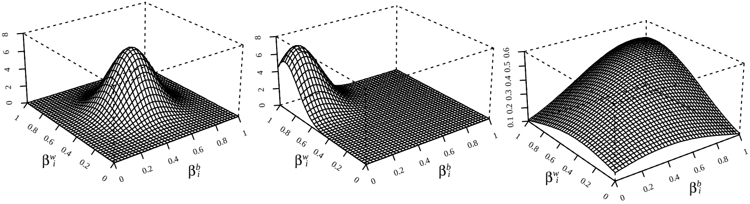
- Moments: $E(Y) = \mu_i$, $V(Y) = \Sigma$, $\text{Cov}(Y_1, Y_2) = \sigma_{12} = \sigma_{21}$.
- $\text{Corr}(Y_1, Y_2) = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$

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.

Key Concepts

Key Concepts

- stochastic and systematic components.

Key Concepts

- stochastic and systematic components.
- simulation as an important tool.

Key Concepts

- stochastic and systematic components.
- simulation as an important tool.
- some basic distributions and what we should know about them.

Key Concepts

- stochastic and systematic components.
- simulation as an important tool.
- some basic distributions and what we should know about them.
- a preview of how these pieces fit together for generalized linear models

Key Concepts

- stochastic and systematic components.
- simulation as an important tool.
- some basic distributions and what we should know about them.
- a preview of how these pieces fit together for generalized linear models

By Next Wednesday: Read *UPM* Chapter 4

Beta (continuous) density

Beta (continuous) density

- Used to model proportions.

Beta (continuous) density

- Used to model proportions.
- We'll use it first to generalize the Binomial distribution

Beta (continuous) density

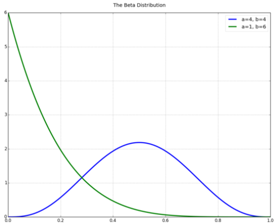
- Used to model proportions.
- We'll use it first to generalize the Binomial distribution
- y falls in the interval [0,1]

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:

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

Standard Parameterization

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

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:

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:

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

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, $\Gamma(x+1) = x! = x(x-1)(x-2)\cdots 1$.

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);`

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?

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:

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:

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

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

Alternative parameterization

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

Alternative parameterization

Set $\mu = E(Y) = \frac{\alpha}{(\alpha+\beta)}$ and $\frac{\mu(1-\mu)\gamma}{(1+\gamma)} = V(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$,

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.

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:

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:

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

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.

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

Beta-Binomial

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

Beta-Binomial

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

How to simulate

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.)

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)

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

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, $j = 1, \dots, N$ (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)

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$(z_j|\tilde{\pi})$

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

Beta-Binomial Analytics

Recall:

Beta-Binomial Analytics

Recall:

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

Beta-Binomial Analytics

Recall:

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

Plan:

Beta-Binomial Analytics

Recall:

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

Plan:

- Derive the joint density of y and π. Then

Beta-Binomial Analytics

Recall:

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

Plan:

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

Beta-Binomial Analytics

Recall:

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

Plan:

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

Beta-Binomial Analytics

Recall:

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

Plan:

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

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

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):

$$\text{BB}(y_i|\mu,\gamma) = \int_0^1 \text{Binomial}(y_i|\pi) \times \text{Beta}(\pi|\mu,\gamma) d\pi$$

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):

$$\begin{aligned} \text{BB}(y_i|\mu,\gamma) &= \int_0^1 \text{Binomial}(y_i|\pi) \times \text{Beta}(\pi|\mu,\gamma) d\pi \\ &= \int_0^1 \mathbb{P}(y_i, \pi|\mu,\gamma) d\pi \end{aligned}$$

Beta-Binomial Analytics

Recall:

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

Plan:

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

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

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

Poisson Distribution

Poisson Distribution

- Begin with an observation period:

Poisson Distribution

- Begin with an observation period:

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.

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

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

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

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
- $\Pr(\text{event at time } t \mid \text{all events up to time } t-1)$ is constant for all t.

Poisson Distribution

- First principles imply:
  - $$\text{Poisson}(y|\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} & \text{for } y_i = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Poisson Distribution

- First principles imply:
  - $$\text{Poisson}(y|\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} & \text{for } y_i = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$
  - $E(Y) = \lambda$

Poisson Distribution

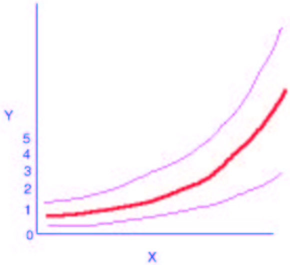
- First principles imply:
  - $$\text{Poisson}(y|\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} & \text{for } y_i = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$
  - $E(Y) = \lambda$
  - $V(Y) = \lambda$

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

- 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.

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.

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

Gamma Density

- Used to model durations and other nonnegative variables

Gamma Density

- Used to model durations and other nonnegative variables
- We'll use first to generalize the Poisson

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.

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

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 $V(Y) = \phi(\sigma^2 - 1)$

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

Negative Binomial

- Same logic as the beta-binomial generalization of the binomial

Negative Binomial

- Same logic as the beta-binomial generalization of the binomial
- Parameters $\phi > 0$ and dispersion parameter $\sigma^2 > 1$

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

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 $V(Y) = \sigma^2\phi$

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: $V(Y) > E(Y)$.

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)

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)

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

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$(\lambda|\phi, \sigma^2)$.

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

Negative Binomial Derivation

Recall:

Negative Binomial Derivation

Recall:

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

Negative Binomial Derivation

Recall:

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

$$\text{NegBin}(y|\phi, \sigma^2) = \int_0^\infty \text{Poisson}(y|\lambda) \times \text{gamma}(\lambda|\phi, \sigma^2) d\lambda$$

Negative Binomial Derivation

Recall:

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

$$\begin{aligned} \text{NegBin}(y|\phi, \sigma^2) &= \int_0^\infty \text{Poisson}(y|\lambda) \times \text{gamma}(\lambda|\phi, \sigma^2) d\lambda \\ &= \int_0^\infty \mathbb{P}(y, \lambda|\phi, \sigma^2) d\lambda \end{aligned}$$

Negative Binomial Derivation

Recall:

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

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