PLSC 502 – Fall 2024
Probability Distributions

October 7, 2024

Why???

Why learn about probability distributions?

- Things we study are both *regular* and *random*.
- Key: **Data-Generating Process**
- Inference: Making claims about the world without getting *all* the data...
- Prediction: Getting things right (on average), but also knowing how precisely we do that.
- For more answer(s) to this question, see this.

What do we need to know?

Distributions are characterized by a combination of their *formulas* and their *moments*...

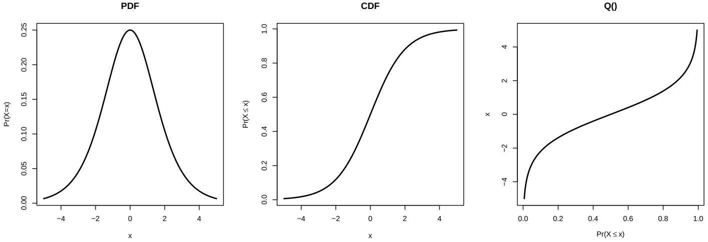
- The *formula* describes the mapping from the outcome space to the probability space.
- The *moments* tell us about the "location," shape, and other characteristics of a particular instance of a distribution.
    - Most distributions are fully characterized by their formula + one or more (usually a small number of) finite moments (often called "parameters").
    - A distribution's *moment generating function* provides a bridge between its moments and the formula for the distribution.

Describing Distributions

Attributes of Distributions

Attribute	a/k/a	Meaning	Function	In R
Density	$f(X)$, "PDF"*	$\Pr(X = x)$	Maps from x to $Pr(X = x)$	`ddist`
Distribution	$F(X)$, "CDF"	$\Pr(X \leq x) = \int_{-\infty}^{x} f(X)dX$	Maps from x to $\Pr(X \leq x)$	`pdist`
Quantile	$Q(p)$, $F_X^{-1}(p)$	Value of x such that $\Pr(X \leq x)$	Maps from $\Pr(X \leq x)$ to x	`qdist`

*Sometimes called a "probability mass function" ("PMF") when X is discrete.



Discrete Distributions

Bernoulli

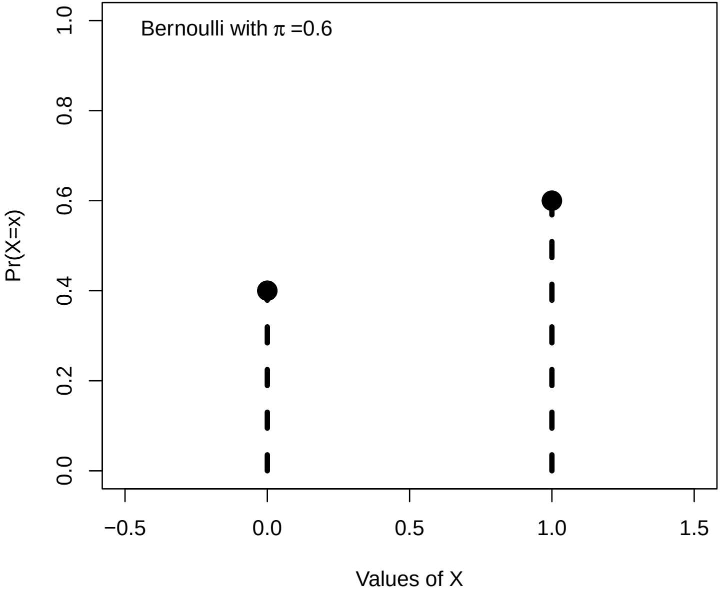
Binary X:

$$X = 0 \text{ with probability } 1 - \pi$$
$$= 1 \text{ with probability } \pi.$$

Density / PDF:

$$f(x) = \begin{cases} 1 - \pi & \text{for } X = 0 \\ \pi & \text{for } X = 1 \end{cases}$$
$$= \pi^x (1 - \pi)^{1-x},\ x \in \{0, 1\}$$

$$X \sim \text{Bernoulli}(\pi)$$


Bernoulli with π =0.6
Pr(X=x)
0.0
0.2
0.4
0.6
0.8
1.0
−0.5
0.0
0.5
1.0
1.5
Values of X

Bernoulli, continued

CDF:

$$
\begin{aligned}
F(x) &= \sum_x f(x) \\
&= \begin{cases} 1-\pi & \text{for } X = 0 \\ 1 & \text{for } X = 1 \end{cases}
\end{aligned}
$$

Expectation:

$$
\begin{aligned}
\mathrm{E}(X) &= \sum_x x f(x) \\
&= (0)(1-\pi) + (1)(\pi) \\
&= \pi
\end{aligned}
$$

Bernoulli, continued

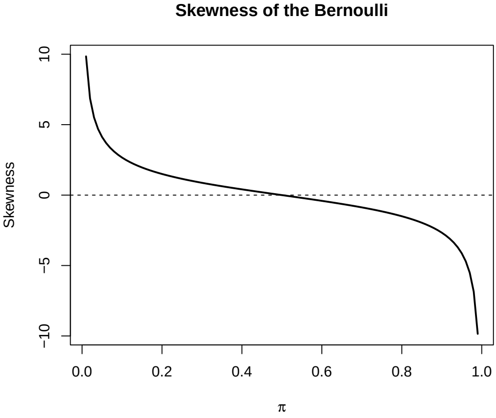
Variance:

$$
\begin{aligned}
\mathrm{Var}(X) &= \sum_x [X - \mathrm{E}(X)]^2 f(x) \\
&= \sum_x [X - \pi]^2 f(x) \\
&= (0 - \pi)^2 (1 - \pi) + (1 - \pi)^2 \pi \\
&= \pi^2 - \pi^3 + \pi - 2\pi^2 + \pi^3 \\
&= \pi - \pi^2 \\
&= \pi(1 - \pi)
\end{aligned}
$$

Even More Bernoulli

Skewness:

$$\text{Skewness} = \frac{(1-\pi)-\pi}{\sqrt{(1-\pi)\pi}}$$



Even More Bernoulli

MGF:

$$\begin{aligned}
\psi(t) &= \int_{-\infty}^{\infty} \exp(tx)\, dF(x) \\
&= \sum_{n=0}^{1} \exp(tn)\pi^n(1-\pi)^{1-n} \\
&= \exp(0)(1-\pi) + \exp(t)\pi \\
&= (1-\pi) + \pi \exp(t)
\end{aligned}$$

Implying:

$$\frac{\partial^k \psi(t)}{\partial^k t} = \pi \exp(t) \ \forall \ k$$

and *raw moments*:

$$\mathrm{E}(X^k) = \pi \ \forall \ k > 0$$

Feel the Bern...

Central moments:

$$M_1 = \pi,$$

$$M_2 = \pi(1-\pi),$$

$$M_3 = \pi(1-\pi)(1-2\pi),$$

etc.

Binomial

Assume n independent binary "trials," each with identical probability of "success" π. Then the number of "successes" in n trials follows a *binomial* distribution:

$$f(x) = \binom{n}{x}\pi^x(1-\pi)^{n-x}$$

where recall that

$$\binom{n}{x} \equiv \frac{n!}{x!(n-x)!}.$$

$$X \sim \text{binomial}(n, \pi).$$

Why "binomial"?

Binomial Theorem:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

e.g.

$$\begin{aligned}
(a+b)^2 &= a^2 + 2ab + b^2, \\
(a+b)^3 &= a^3 + 3a^2 b + 3ab^2 + y^3, \\
(a+b)^4 &= a^4 + 4a^3 b + 6a^2 b^2 + 4ab^3 + b^4,
\end{aligned}$$

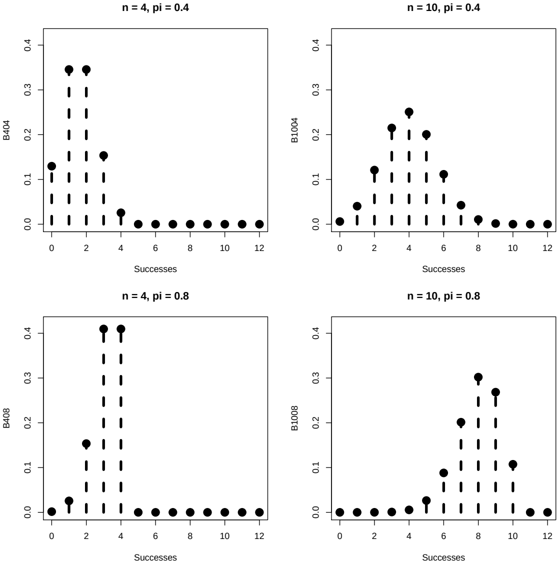
Binomial: Derivation

For $n = 2$:

$$\begin{aligned}
\Pr(X = 0) &= \Pr(X_1 = 0, X_2 = 0) \\
&= \Pr(X_1 = 0) \times \Pr(X_2 = 0) \\
&= (1 - \pi)^2
\end{aligned}$$

$$\begin{aligned}
\Pr(X = 1) &= \Pr(X_1 = 1, X_2 = 0 \text{ or } X_1 = 0, X_2 = 1) \\
&= \Pr(X_1 = 1) \times \Pr(X_2 = 0) + \Pr(X_1 = 0) \times \Pr(X_2 = 1) \\
&= \pi(1 - \pi) + (1 - \pi)\pi \\
&= 2[\pi(1 - \pi)]
\end{aligned}$$

$$\begin{aligned}
\Pr(X = 2) &= \Pr(X_1 = 1, X_2 = 1) \\
&= \Pr(X_1 = 1) \times \Pr(X_2 = 1) \\
&= \pi^2
\end{aligned}$$


n = 4, pi = 0.4
n = 10, pi = 0.4
n = 4, pi = 0.8
n = 10, pi = 0.8
B404
B1004
B408
B1008
Successes
0.0
0.1
0.2
0.3
0.4
0 2 4 6 8 10 12

CDF:

$$
\begin{aligned}
F(x) &= \sum_{x} f(x) \\
&= \sum_{j=0}^{x} \binom{n}{j} \pi^{j} (1-\pi)^{n-j}
\end{aligned}
$$

Expectation:

$$
\mathrm{E}(X) = n\pi,
$$

Variance:

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_x [X - \mathrm{E}(X)]^2 f(x) \\
&= \sum_x (X - \pi n)^2 \binom{n}{x} \pi^x (1-\pi)^{n-x} \\
&= n\pi(1-\pi).
\end{aligned}$$

Skewness:

$$\text{Skewness} = \frac{1 - 2\pi}{\sqrt{n\pi(1-\pi)}}$$

The Binomial...

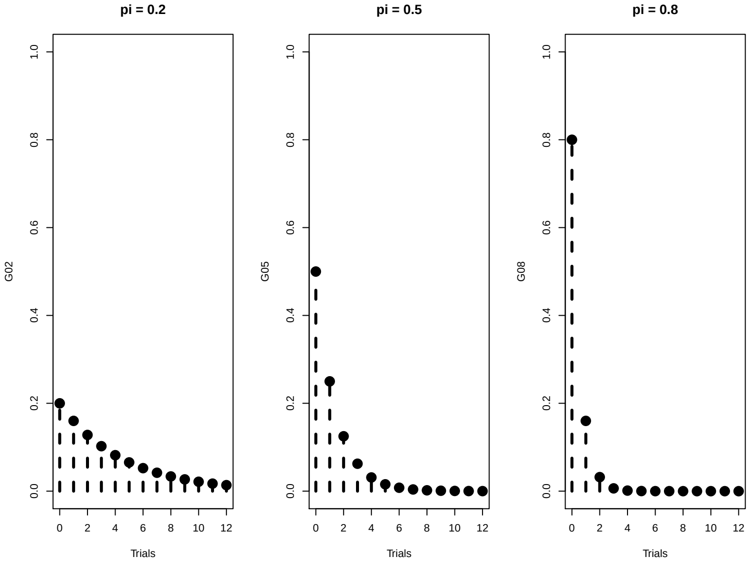
- Is *unimodal* (except in certain cases),
- has median $\lceil n\pi \rceil$ or $\lfloor n\pi \rfloor$,
- has mode $\lceil (n+1)\pi \rceil$ or $\lfloor (n+1)\pi \rfloor$,
- has skewness that is:
  - zero when $\pi = 0.5$
  - positive when $\pi < 0.5$
  - negative when $\pi > 0.5$

The number of independent Bernoulli trials needed to achieve *one* success is a *geometric* random variable.

PDF:

$$f(x) = \pi(1-\pi)^{x-1}$$

$$X \sim \text{geometric}(\pi).$$



pi = 0.2
pi = 0.5
pi = 0.8
G02
G05
G08
Trials
Trials
Trials
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
12

CDF:

$$\begin{aligned} F(x) &= \sum_{j=1}^{x} \pi(1-\pi)^{x-1} \\ &= 1-(1-\pi)^{x} \end{aligned}$$

Expectation:

$$\mathrm{E}(X) = \frac{1}{\pi}$$

Variance:

$$\mathrm{Var}(X) = \frac{1-\pi}{\pi^2}$$

Negative Binomial

The number of *failures we observe* (x) before achieving the rth success in a series of independent Bernoulli trials (each with equal probability of success π) is distributed according to a *negative binomial* distribution.

PDF:

$$f(x) = \binom{r+x-1}{r-1}\pi^r(1-\pi)^x$$

CDF:

$$
\begin{aligned}
F(x) &= \sum_{j=0}^{x} \binom{r+j-1}{r-1} \pi^r (1-\pi)^j \\
&= 1 - \mathrm{CDF}_{\mathrm{binomial}}
\end{aligned}
$$

Expected value:

$$
\mathrm{E}(X) = \frac{(1-\pi)r}{\pi}
$$

Even More Negative Binomial

Variance:

$$\mathsf{Var}(X) = \frac{(1-\pi)r}{\pi^2}.$$

Skewness:

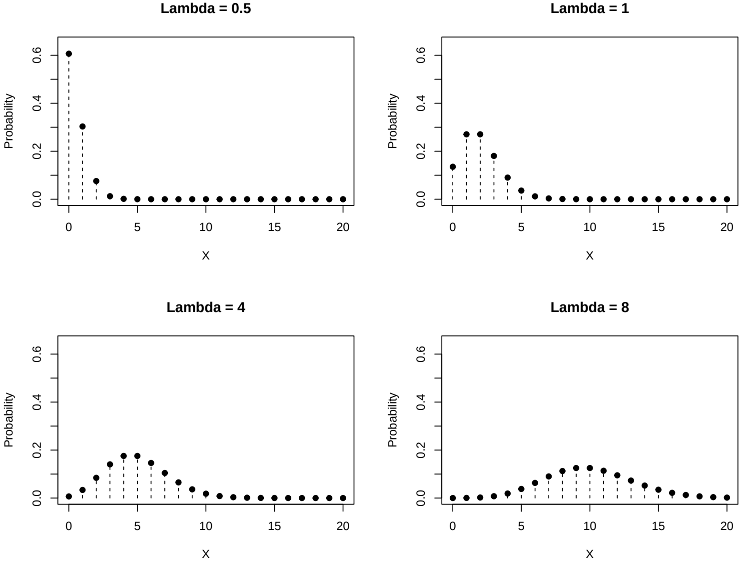
$$\mathsf{Skewness} = \frac{1+\pi}{\sqrt{\pi r}}$$

Poisson

For n independent Bernoulli trials with (sufficiently small) probability of success π and where $n\pi \equiv \lambda > 0$, the probability of observing exactly x total "successes" as the number of trials grows without limit is the *Poisson distribution*.

PDF:

$$
\begin{aligned}
f(x) &= \lim_{n \to \infty} \left[\binom{n}{x} \left(\frac{\lambda}{n}\right)^{x} \left(1 - \frac{\lambda}{n}\right)^{n-x} \right] \\
&= \frac{\lambda^{x} \exp(-\lambda)}{x!}.
\end{aligned}
$$


Lambda = 0.5
Lambda = 1
Lambda = 4
Lambda = 8
Probability
X
0.0
0.2
0.4
0.6
0
5
10
15
20

More Poisson

CDF:

$$F(x) = \sum_{j=0}^{x} \frac{\lambda^j \exp(-\lambda)}{x!}.$$

Mean & Variance:

$$E(X) = \mathrm{Var}(X) = \lambda$$

All higher moments are zero...

Alternative Poisson

Independent, *constant-probability* events occurring in time...

$$\text{“Arrival rate”} = \lambda$$

Implies:

$$\begin{aligned} \Pr(\text{Event in } (t, t+h]) &= \lambda \\ \Pr(\text{No event in } (t, t+h]) &= 1 - \lambda \end{aligned}$$

$$N_{\text{Events occurring in } (t,t+h]} = \frac{\exp(-\lambda h)\lambda h^x}{x!}$$

If $h = 1 \;\forall\; h$, then:

$$f(x) = \frac{\exp(-\lambda)\lambda^x}{x!}$$

Multinomial

Imagine K possible distinct *outcomes* for each "trial," where each possible outcome $k \in \{1, 2, ...K\}$ has π_k and $\sum_{k=1}^{K} \pi_k = 1$.

Define x_k = number of times we observe outcome k out of n trials.

Then for:

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{pmatrix}$$

the PDF is:

$$f(\mathbf{x}) = \frac{n!}{x_1! x_2! ... x_K!} \pi_1^{x_1} \pi_2^{x_2} ... \pi_K^{x_K}$$

Multinomial, continued

Expected value:

$$E(\mathbf{X}) \equiv E\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{pmatrix} = n\begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_K \end{pmatrix}$$

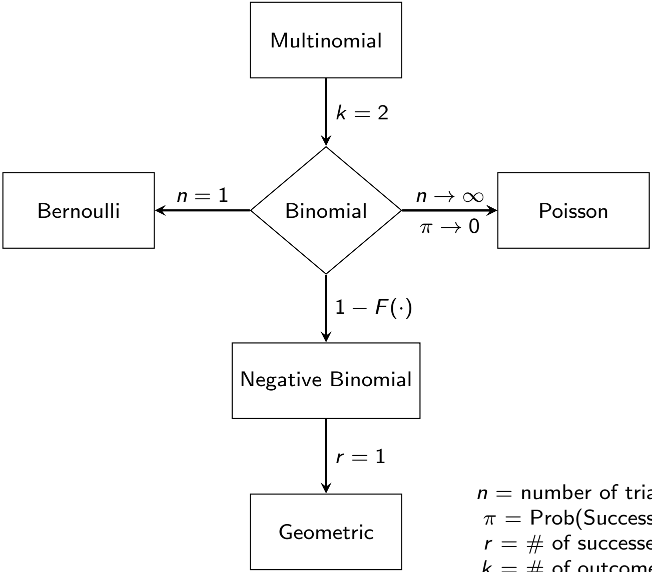
Variance:

$$\mathrm{Var}(\mathbf{X}) \equiv \mathrm{Var}\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{pmatrix} = n\begin{bmatrix} \pi_1(1-\pi_1) \\ \pi_2(1-\pi_2) \\ \vdots \\ \pi_K(1-\pi_K) \end{bmatrix}$$

Covariance between X_s and X_t, $s \neq t$:

$$\mathrm{Cov}(X_s, X_t) = -n\pi_s\pi_t$$

Schematic



$n =$ number of trials
$\pi =$ Prob(Success)
$r = \#$ of successes
$k = \#$ of outcomes

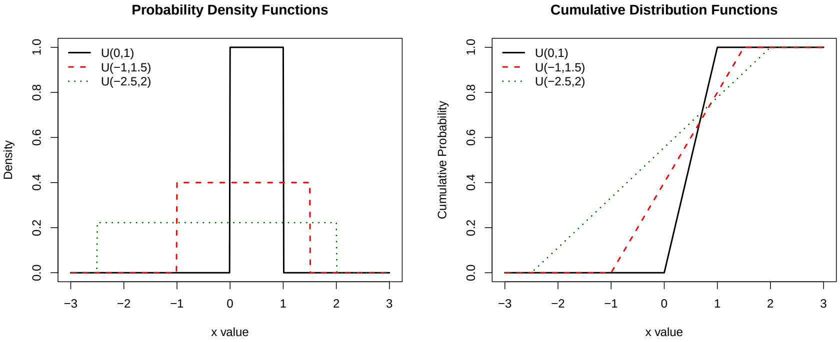
Continuous Distributions

The Uniform Distribution

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{for } x < a \text{ or } x > b. \end{cases}$$

$$F(x) = \int f(x)dx = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \le x < b \\ 1 & \text{for } x \ge b \end{cases}$$

Uniform PDFs and CDFs



Uniform Characteristics

Expected value / "mean":

$$\mathrm{E}(X) = \check{X} = \frac{a+b}{2}$$

Mode:

$$\mathrm{mode}(X) = [a, b]$$

Variance:

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

Skewness:

$$\mathrm{Skewness}(X) = 0$$

The Standard Uniform Distribution

Special case of the uniform:

$$X \sim U(0, 1)$$

It has the property that:

$$X \sim 1 - X \sim U(0, 1).$$

In addition, the CDF is also unique:

$$F(x) = x$$

The Normal Distribution

The Normal density is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

abbreviated:

$$X \sim \phi_{\mu,\sigma^2}.$$

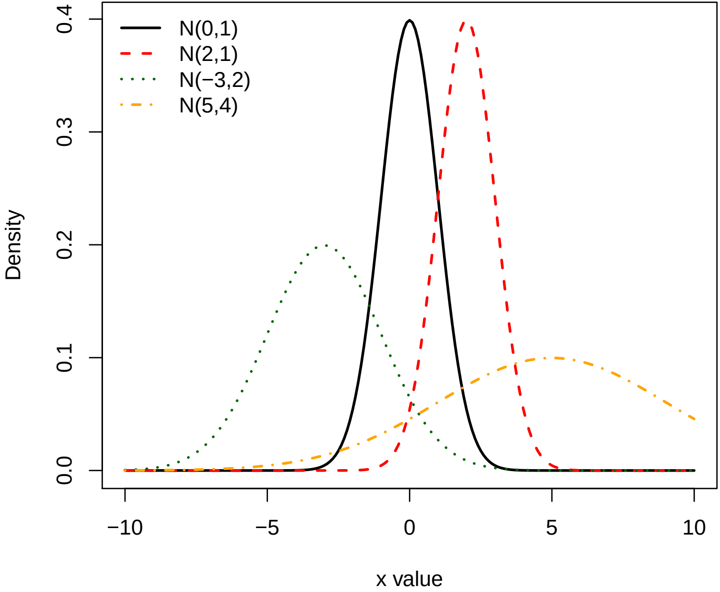
The CDF is:

$$\begin{aligned} F(x) &= \Phi_{\mu,\sigma^2}(x) \\ &= \frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right], \end{aligned}$$

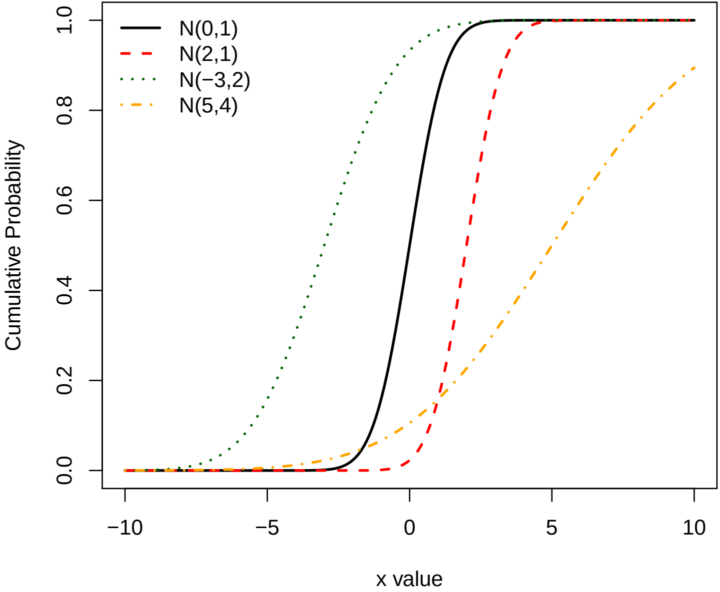
where

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt.$$

Various Normal Densities



Various Normal CDFs



Why Normal?

For $i = \{1, 2, \ldots N\}$ i.i.d. X_i with $\mu_i < \infty$ and $\sigma_i^2 > 0$, define:

$$X = \sum_{i=1}^{N} X_i.$$

Then

$$\begin{aligned} \mathrm{E}(X) &= \sum_{i=1}^{N} \mu_i \\ &= \mu < \infty \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}(X) &= \sum_{i=1}^{N} \sigma_i^2 \\ &= \sigma^2 < \infty. \end{aligned}$$

Central Limit Theorem

The "CLT":

$$\lim_{N\to\infty} X = \lim_{N\to\infty} \sum_{i=1}^{N} X_i \xrightarrow{D} N(\cdot)$$

"...we often think of a normal distribution as being appropriate when the observed variable X can take on a range of continuous values, and when the observed value of X can be thought of as the product of a large number of relatively small, independent "shocks" or perturbations."

Properties of the Normal Distribution

The normal is a <u>two-parameter distribution</u>, where

$$\mu \in (-\infty, \infty)$$

and

$$\sigma^2 \in (0, \infty).$$

For $X \sim N(\mu, \sigma^2)$:

- X has support in $\mathfrak{R}$
- $\text{Skewness}(X) = 0$
- X is *mesokurtic*
- If $X \sim N(\mu, \sigma^2)$, then $aX + b \sim N(a\mu + b, a^2\sigma^2)$

The Standard Normal Distribution

Linear transformation:

- $b = \frac{-\mu}{\sigma}$,
- $a = \frac{1}{\sigma}$.

Yields:

$$\begin{aligned} ax + b &\sim N(a\mu + b, a^2\sigma^2) \\ &\sim N(0, 1) \end{aligned}$$

In other words:

- If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{(x-\mu)}{\sigma} \sim N(0, 1)$.
- The PDF is:

$$f(z) \equiv \phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(z)^2}{2}\right]$$

Similarly, we often write the CDF for the standard normal as $\Phi(\cdot)$.

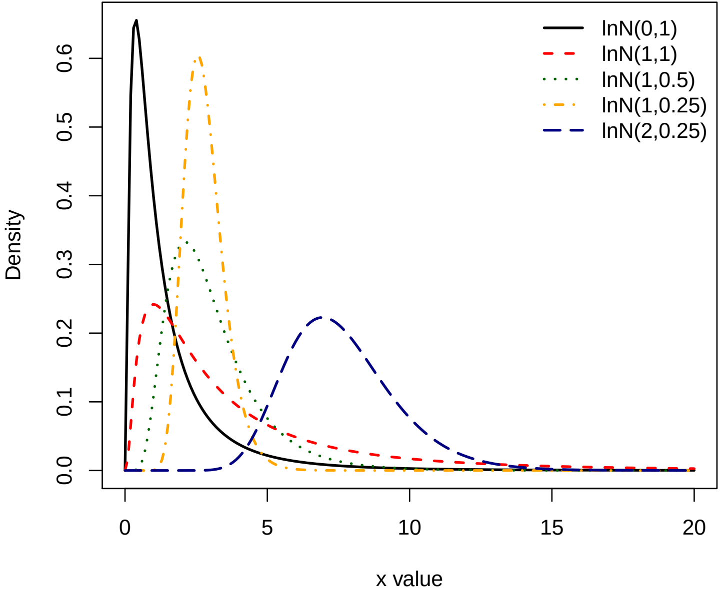
The Log-Normal Distribution

$$Y = \exp(X) \sim \mathrm{LogN}(m, s^2)$$

PDF:

$$f(y) = \frac{1}{ys\sqrt{2\pi}} \times \exp\left[\frac{-(\ln y - m)^2}{2s^2}\right].$$

Log-Normal PDFs



The χ^2 Distribution

PDF:

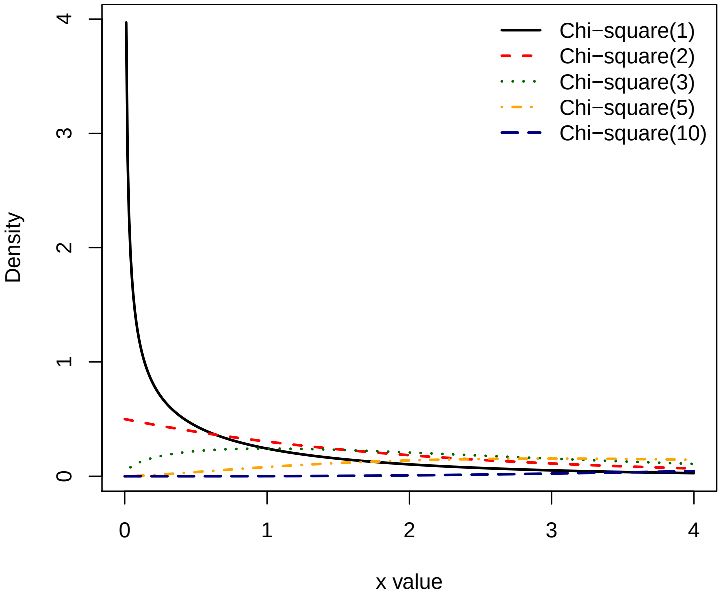
$$\begin{aligned} f(w) &= \frac{1}{2^k\Gamma(k)} w^k \exp\left[\frac{-w}{2}\right] \\ &= \frac{w^{\frac{k-2}{2}}\exp(\frac{-w}{2})}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} \end{aligned}$$

where $\Gamma(k) = \int_0^\infty t^{k-1}\exp(-t)\,dt$.

CDF:

$$F(w) = \frac{\gamma(k/2, w/2)}{\Gamma(k/2)}$$

χ^2 Densities



χ^2 *Characteristics*

For $W \sim \chi^2_k$:

$$E(W) = k$$

and:

$$\mathrm{Var}(W) = 2k.$$

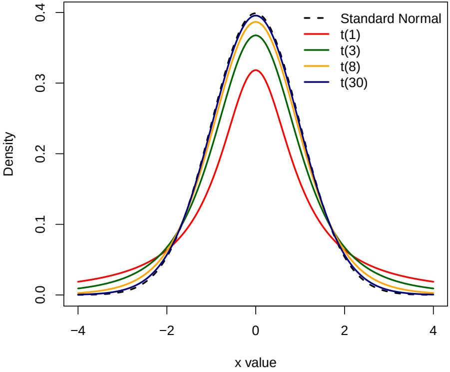
Also:

- If $Z \sim N(0,1)$, then $Z^2 \sim \chi^2_1$
- If $W_j \sim \chi^2_j$ and $W_k \sim \chi^2_k$ and independent, then
  - $W_j + W_k$ is $\sim \chi^2_{j+k}$ and more generally
  - $\sum_{i=1}^{k} W_i \sim \chi^2_k$.

Student's t Distribution

PDF:

$$f(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\,\Gamma(\frac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-(\frac{k+1}{2})}$$



Student's t Characteristics

- Mean: $\mu = 0$
- Variance: $\sigma^2 = \frac{k}{k-2}$
- Skewness is 0 for $k > 3$; otherwise undefined
- $t_k \to N(0,1)$ as $k \to \infty$
- If if $Z \sim N(0,1)$, $W \sim \chi^2_k$, and Z and W are independent, then

$$\frac{Z}{\sqrt{W/k}} \sim t_k$$

and

$$\frac{Z^2}{W/k} \sim t_k.$$

The F Distribution

PDF:

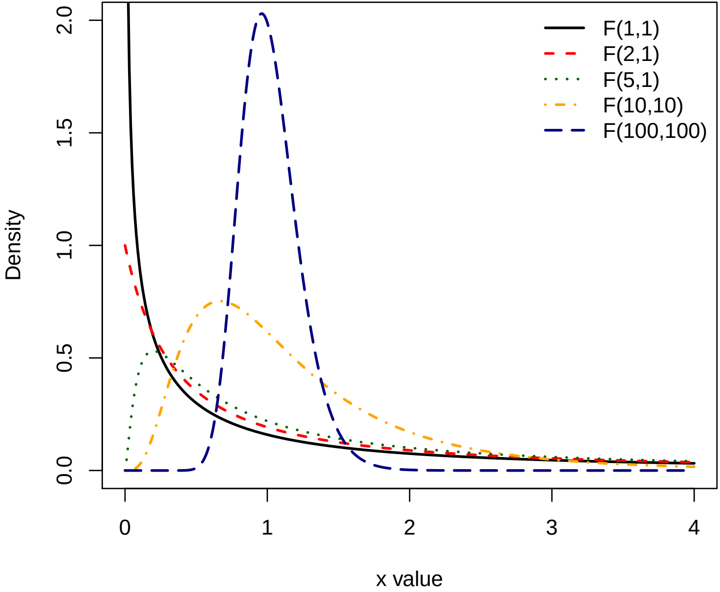
$$f(x) = \frac{\left(\frac{k\,x}{k\,x+\ell}\right)^{k/2}\left(1-\frac{k\,x}{k\,x+\ell}\right)^{\ell/2}}{x\,\mathrm{B}(k/2,\ell/2)}$$

where $\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt$.

And we say:

$$X \sim F_{k,\ell}$$

F Densities



F Characteristics

For an $F_{k,\ell}$ distribution:

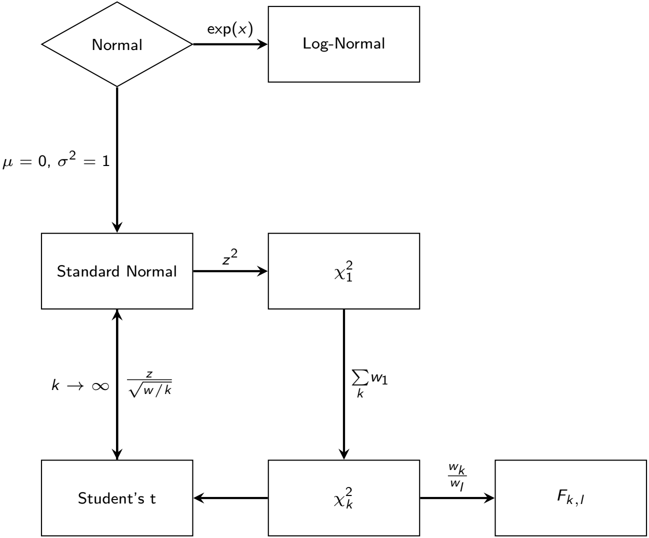
- $\mathrm{E}(X) = \frac{\ell}{\ell-2}$
- $\mathrm{Var}(X) = \frac{2\,\ell^2\,(k+\ell-2)}{k(\ell-2)^2(\ell-4)}$
- Skewness $= \frac{(2k+\ell-2)\sqrt{8(\ell-4)}}{(\ell-6)\sqrt{k(k+\ell-2)}}$
- For independent $W_1 \sim \chi^2_k$ and $W_2 \sim \chi^2_\ell$:

$$\frac{W_1}{W_2} \sim F_{k,\ell}$$

Relationships Among Continuous Distributions

<u>Variables</u>

$$\begin{aligned} x &\sim N(\mu, \sigma^2) \\ z &\sim N(0,1) \\ w_k &\sim \chi^2_k \end{aligned}$$



Summary

A Few Distributions

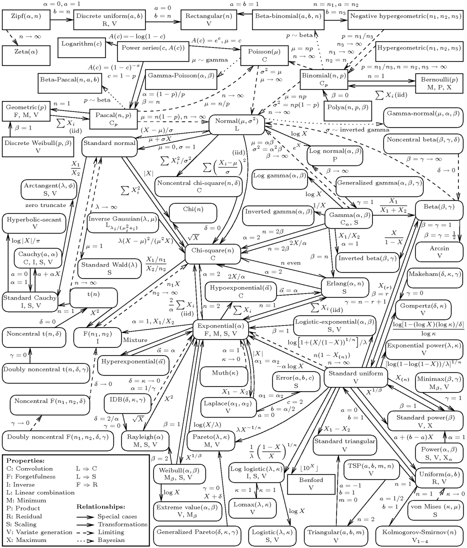
Distribution	First Parameter	Meaning	Second Parameter	Meaning
Bernoulli	π	Pr(Success)		
Binomial	n	N of trials	π	Pr(Success)
Geometric	π	Pr(Success)		
Negative Binomial	r	Rank of "success"	π	Pr(Success)
Poisson	λ	"Arrival rate"		
Uniform	a	Minimum	b	Maximum
Normal	μ	Mean	σ^2	Variance
Lognormal	m	Mean	s^2	Variance
Student's t	k	Degrees of freedom		
Chi-Square	k	Degrees of freedom		
F	k	Degrees of freedom	ℓ	Degrees of freedom

Other Useful Distributions

There are a lot; here, in no particular order:

- The beta distribution – a continuous two-parameter distribution defined on $[0, 1]$; widely used to model percentages / proportions, and as a "prior" in Bayesian analysis.
- The exponential distribution – a continuous one-parameter distribution defined on $[0, \infty)$ that is related to the Poisson distribution (and others).
- The Weibull distribution – a continuous two parameter distribution defined on $[0, \infty)$ widely used in survival analysis.
- The gamma distribution – a continuous two parameter distribution defined on $[0, \infty)$ that encompasses the exponential, chi-square, and other distributions as special cases.
- The logistic distribution – a continuous two parameter distribution defined on $\mathbb{R}$ that resembles the Normal and is widely used in statistics (e.g., logistic regression) and machine learning.
- The Cauchy distribution – a continuous two parameter distribution defined on $\mathbb{R}$ and that is interesting mainly because it has no defined MGF.

(Almost) All The Distributions



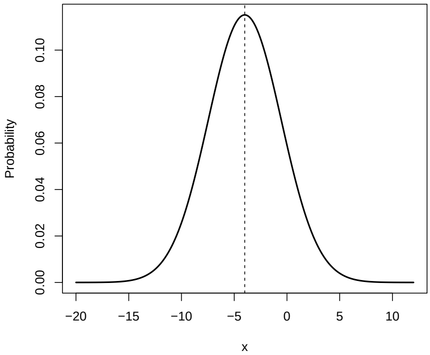
(Source)

Practical Things

How To Plot A Density / PDF

"Plot a Normal distribution with a mean of -4 and a variance of 12."

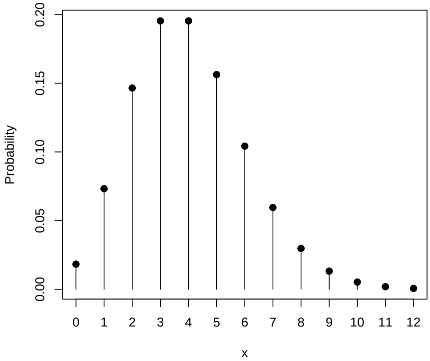
```
> x<-seq(-20,12,by=0.1)
> PlotNorm<-dnorm(x,-4,sqrt(12))
```



Plotting A Discrete Density

"Plot a Poisson distribution with $\lambda = 4$."

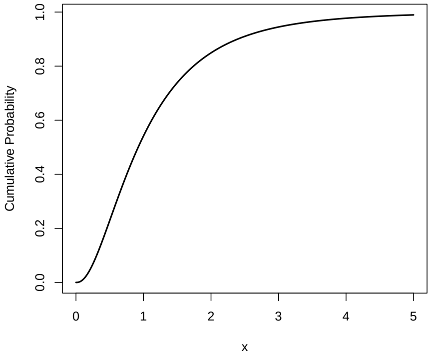
```
> x<-seq(0,12)
> PlotPois<-dpois(x,4)
```



How To Plot A CDF

"Plot the CDF of an F distribution with $k = 5$ and $\ell = 12$."

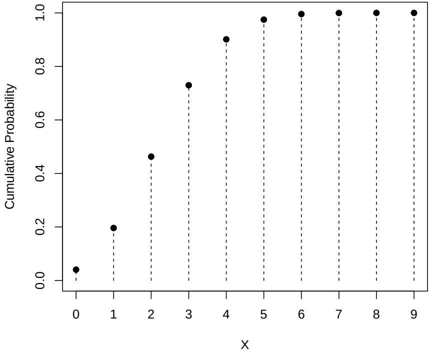
```
> x<-seq(0,5,by=0.01)
> PlotF<-pf(x,5,12)
```



CDF: Discrete Distribution

"Plot the CDF of a binomial distribution with $\pi = 0.3$ and $n = 9$."

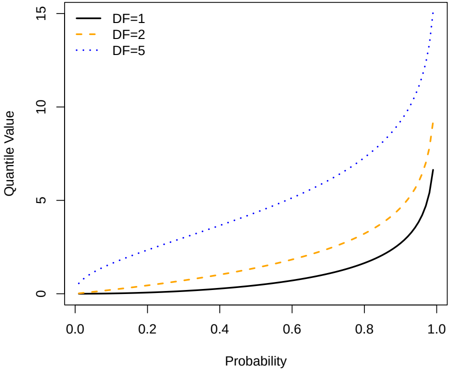
```
> x<-seq(0,9)
> PlotBinom9<-pbinom(x,9,0.3)
```



Quantiles of a Distribution

"Plot the quantiles of three χ^2 distributions with one, two, and five degrees of freedom."

```
> P<-seq(0.01,0.99,by=0.01) # probabilities
> ChiSq1<-qchisq(P,1) # df=1
> ChiSq2<-qchisq(P,2) # df=2
> ChiSq5<-qchisq(P,5) # df=5
```



Simulating Random Variables

Commands for Generating Random Variates

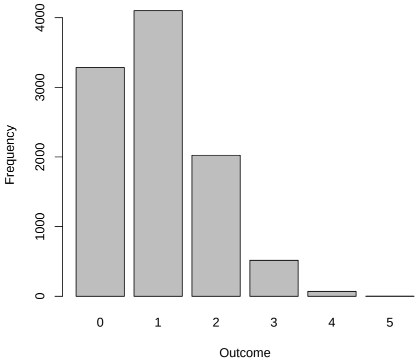
Distribution	R	Stata
Binomial(n, π)	`rbinom()`	`rbinomial`
Geometric(π)	`rgeom()`	("by hand")
Negative Binomial(n, π)	`rnbinom()`	`rnbinomial`
Poisson(λ)	`rpois()`	`rpoisson`
Uniform(0, 1)	`runif()`	`runiform`
Normal(0, 1)	`rnorm()`	`rnormal`
Lognormal(0, 1)	`rlnorm()`	`xlgn`*
Student's $t(k)$	`rt()`	`rt`
Chi-Square(k)	`rchisq()`	`rchi2`
$F(k, \ell)$	`rf()`	`rndf`*

Note: Stata commands marked with an asterisk are from Hilbe's `-rnd-` group of commands.

Drawing From A Distribution

"Draw 10,000 random draws from a binomial distribution with $n = 5$ and $\pi = 0.2$."

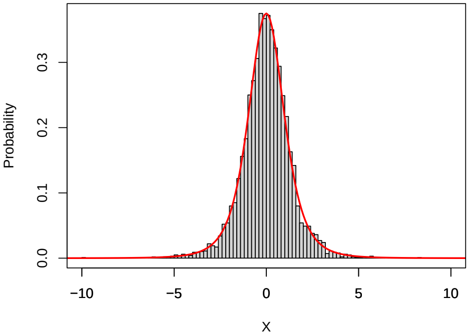
```
> Xbinom5point2<-rbinom(10000,5,0.2)
```



More Draws From A Distribution

"Draw 5000 random draws from a *t* distribution with 4 degrees of freedom, and compare the distribution of values to the theoretical density."

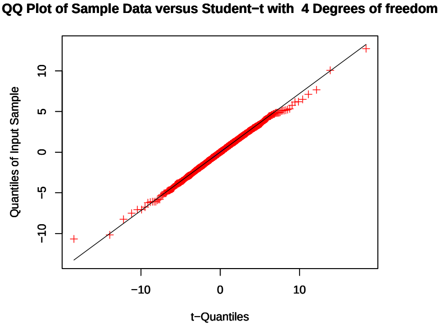
```
> TDraws<-rt(5000,4)
```



(continued...)

(Same comparison, using a Q-Q plot...)

```
> library(Dowd)
> TQQPlot(TDraws,4)
```

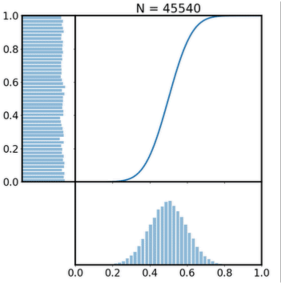


Generating Random Variates

One way to generate random variables is to use *inverse transform sampling*. The intuition:

- Since a probability distribution is a function that maps from the domain of X to the (probability) range [0,1]...
- ...we can do the "reverse mapping" from [0,1] to the domain of X.
- In $U[0,1]$, any given probability of a possible value of X is equal, so...
- ...sampling from $U[0,1]$ and transforming the result by the *generalized inverse* of $F(X)$ is the same as drawing from $f(X)$

Click on this image for an animation of the intuition:



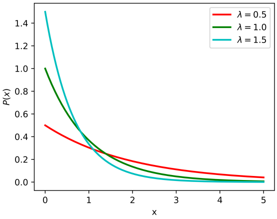
Example: The Exponential Distribution

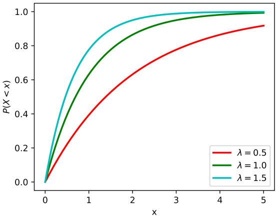
The *exponential distribution* has density:

$$f(x) = \lambda \exp(-\lambda x)$$

for $x \geq 0$ and $\lambda > 0$, with CDF:

$$F(x) = 1 - \exp(-\lambda x).$$





Exponential Distribution (continued)

If we solve $= 1 - \exp(-\lambda x)$ for x with $y \in [0, 1]$, we get:

$$x = -\left(\frac{1}{\lambda}\right) \ln(1 - U)$$

<u>So we can</u>:

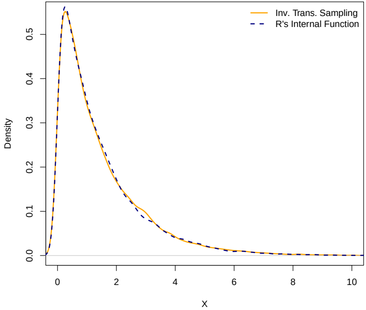
1. Set a value $\lambda = k$
2. Draw a value U_i from $U(0, 1)$
3. Calculate $X_i = -\left(\frac{1}{\lambda}\right) \ln(1 - U_i)$
4. Know that X_i is a random draw from an exponential distribution with $\lambda = k$

Exponential Example

Some code:

```
> nreps<-20000
> lambda<-0.7
>
> set.seed(7222009)
> U<-runif(nreps,0,1)
> XITS<- (-((1)/(lambda))) * log(1-U) # Inv. trans. sampling
> set.seed(7222009)
> XR<-rexp(nreps,rate=0.7)            # R's internal function
```

The result:



Pseudo-Random Numbers and "Seeds"

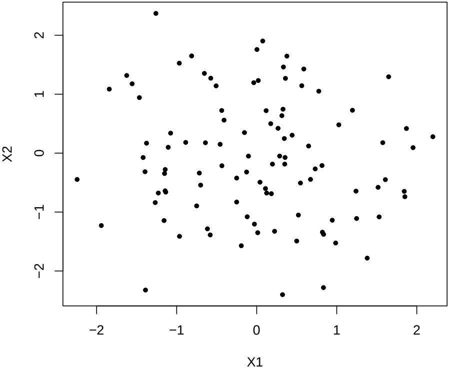
```
> seed<-3229      # calling "seed" some thing
> set.seed(seed) # setting the system seed
> rt(3,1)        # three draws from a t distrib. w/1 d.f.
[1] -0.1113 -0.7306  1.9839

> seed<-1077
> set.seed(seed) # resetting the seed
> rt(3,1)        # different values for the draws
[1]   -0.5211    7.9161 -155.3186

> seed<-3229        # original seed
> set.seed(seed)
> rt(3,1)           # identical values of the draws
[1] -0.1113 -0.7306  1.9839
```

The right way:

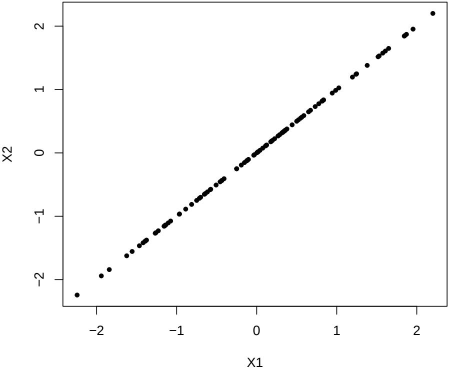
```
> X<-matrix(NA,nrow=100,ncol=2)
> set.seed(7222009)
> for(i in 1:2){
+   X[,i]<-rnorm(100)}
```



Seeds and Simulations (continued)

The wrong way:

```
> X<-matrix(NA,nrow=100,ncol=2)
> for(i in 1:2){
+   set.seed(7222009)
+   X[,i]<-rnorm(100)}
```



Best Practices

In general:

1. **Always** set a seed value.
2. Try to use a single/consistent seed value over time/projects.
3. Keep seeds *outside* loops/`apply` statements.
6. Use `ddists` and `pdists` for theoretical quantities, `rdists` for generating simulated "data" / variates.
5. Plot discrete distributions discretely....
6. ...and continuous distributions continuously.