Goodness of Fit Tests

SYS-611: Simulation and Modeling

Paul T. Grogan, Ph.D.
Assistant Professor
School of Systems and Enterprises

Agenda

1. Review of Process Generators
2. Goodness of Fit Tests for Discrete Data
3. Goodness of Fit Tests for Continuous Data
4. Engineering Model Validation

Readings: J.V. Farr, “Review of Probability and Statistics,” Ch. 3 in *Simulation of Complex Systems and Enterprises,* Stevens Institute of Technology, 2007, pp. 39-47.

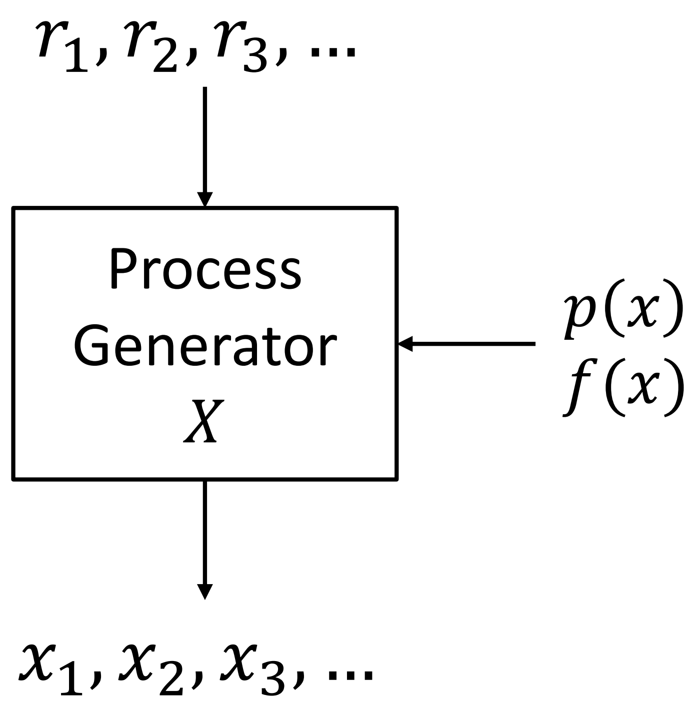
Optional: S.M. Ross, “Statistical Validation Techniques,” Ch. 9 in *Simulation*, McGraw-Hill, 2006, pp. 219-229.

G. Hazelrigg, “Thoughts on Model Validation for Engineering Design,” *Proceedings of the DETC’03*, Chicago, IL, 2003.

Review of Process Generators

Process Generator

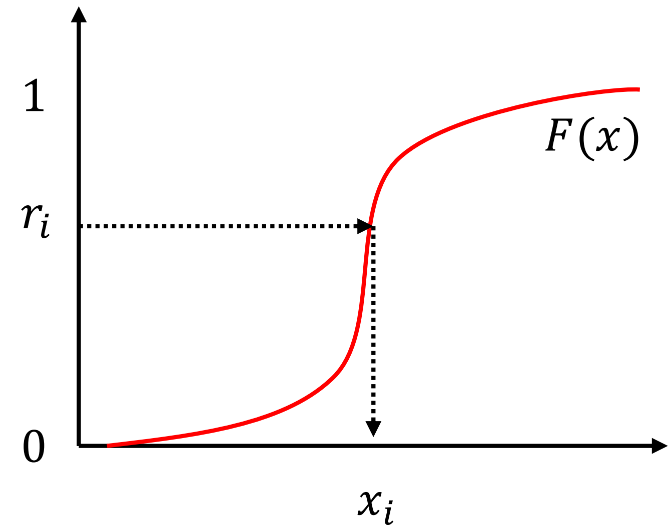
- Method to generate *simulated* random variables
- Requires a known probability distribution (PMF/PDF)
- Two classes of methods:
  - Inverse Transform
  - Accept-Reject



Inverse Transform Method

- Requires a known cumulative distribution function (CDF)
  - Must be able to describe the "inverse"
  - Visualized by tracing random y-axis to CDF

$$r = F(x) \Rightarrow x = F^{-1}(r)$$



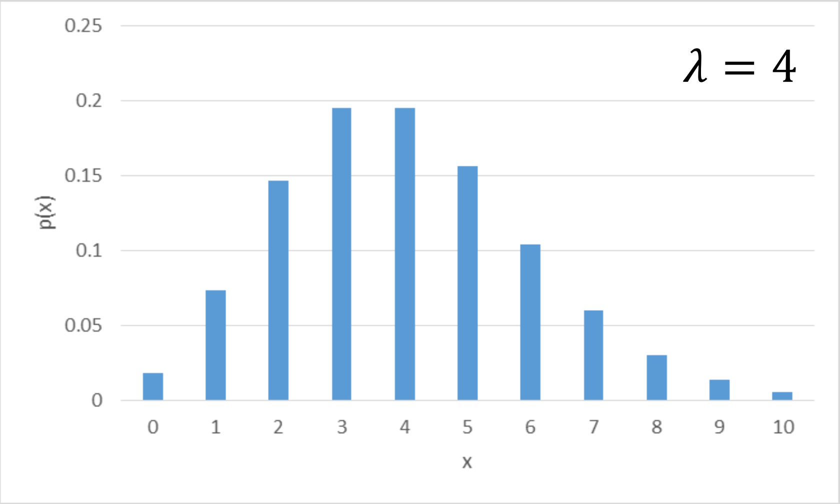
Poisson Generator

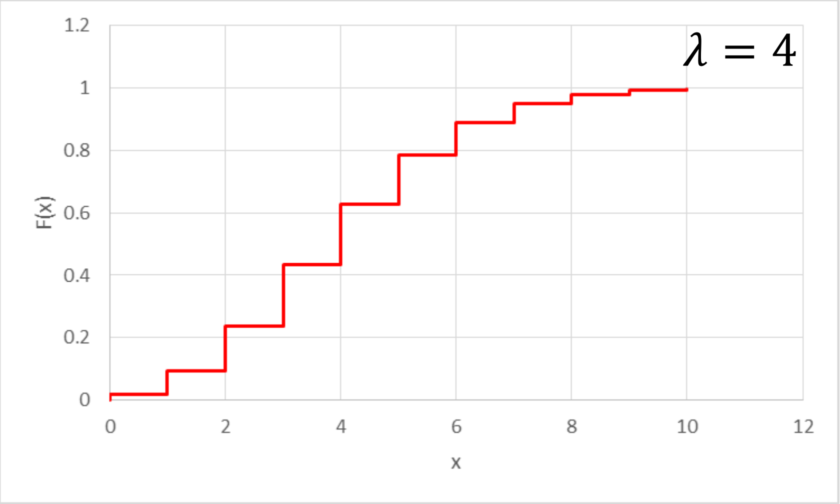
λ: mean event rate

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$F(x) = \sum_{i=0}^{x} \frac{\lambda^i}{i!} e^{-\lambda}$$

$$\Rightarrow x = \begin{cases} 0 & F(x) \le e^{-\lambda} \\ 1 & e^{-\lambda} < F(x) \le \sum_{i=0}^{1} \frac{\lambda^i}{i!} e^{-\lambda} \\ 2 & \sum_{i=0}^{1} \frac{\lambda^i}{i!} e^{-\lambda} < F(x) \le \sum_{i=0}^{2} \frac{\lambda^i}{i!} e^{-\lambda} \\ & \vdots \end{cases}$$





Exponential Generator

$\frac{1}{\lambda}$: mean inter-event time

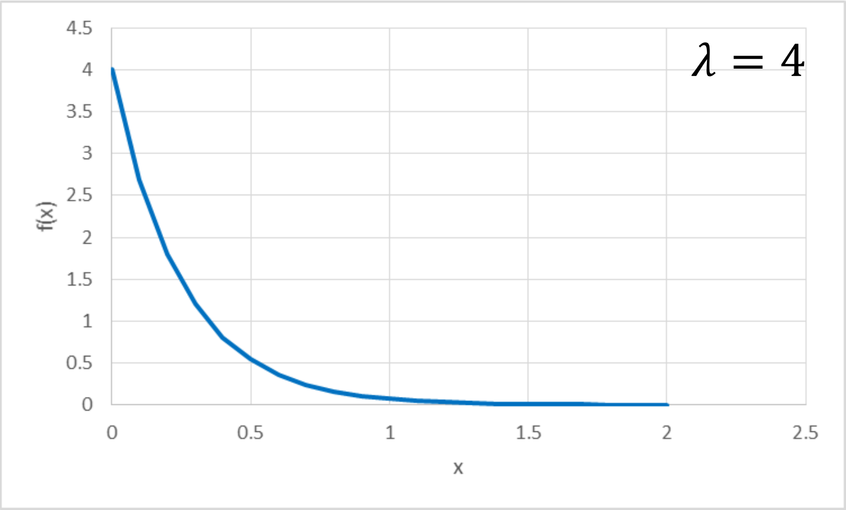
λ: mean event rate

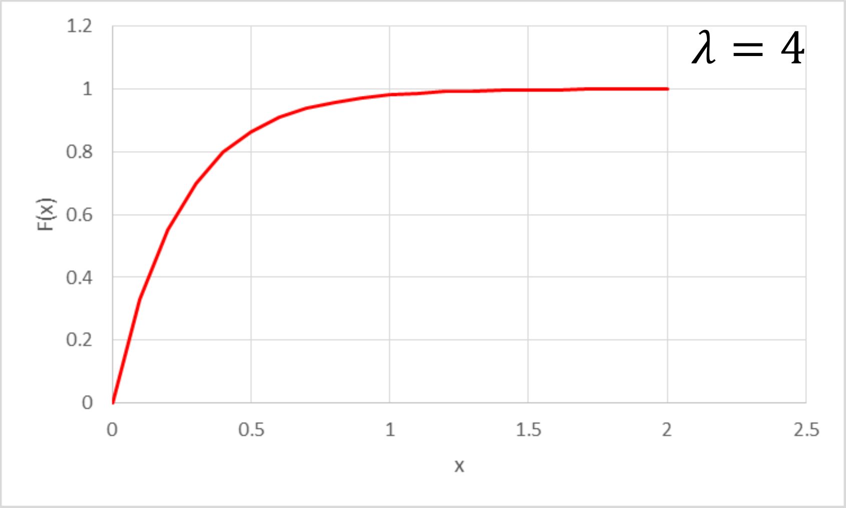
$$f(x) = \lambda e^{-\lambda x}$$

$$F(x) = \int_{i=0}^{x} \lambda e^{-\lambda i} di$$

$$= 1 - e^{-\lambda x}$$

$$\Rightarrow x = \frac{-\ln(1 - F(x))}{\lambda}$$





Goodness of Fit Tests for Discrete Data

Discrete Goodness of Fit Tests

A **Goodness of Fit** test checks if samples observed of a random variable come from a particular distribution

- Validate process generators with commonly-used distributions (e.g. Poisson, Binomial)
- **Pearson's Chi-squared (χ^2) Test**: requires large (>5) samples observed from most values
- **Fisher's Exact Test** (among others): handles small sample sizes but more complex

Pearson's Chi-squared (χ^2) Test

- **Pearson's chi-squared test** evaluates whether a set of samples follow a hypothesized distribution

  H_0: the data are consistent with a specified distribution

  H_a: the data are not consistent with a specified distribution

  - Should have 5+ observations for at least 80% of values
- Compares the following:
  - *Expected* or *theoretical* frequencies of categories
  - *Observed* or *actual* frequencies of categories

Dice Roller Data Set

- X: number of {3, 4, 5, 6} rolled in 10 dice
- Observed data (123 samples):

x_i	0	1	2	3	4	5	6	7	8	9	10
O_i	0	0	3	1	6	25	26	25	23	11	3

- Expected if $X \sim \text{binomial}(n = 10, p = 2/3)$:

$$E_i = N \cdot p(x_i), \qquad p(x) = \binom{10}{x}\left(1 - \frac{2}{3}\right)^{10-x}\left(\frac{2}{3}\right)^{x}$$

x_i	0	1	2	3	4	5	6	7	8	9	10
E_i	0.0	0.0	0.4	2.0	7.0	16.8	28.0	32.0	24.0	10.7	2.1

Combine for 5+ expected observations

Dice Roller Data Set

- X: number of {3, 4, 5, 6} rolled in 10 dice
- Observed data (123 samples):

x_i	0-4	5	6	7	8	9	10
O_i	10	25	26	25	23	11	3

- Expected if $X \sim \text{binomial}(n = 10, p = 2/3)$:

$$E_i = N \cdot p(x_i), \qquad p(x) = \binom{10}{x}\left(1 - \frac{2}{3}\right)^{10-x}\left(\frac{2}{3}\right)^{x}$$

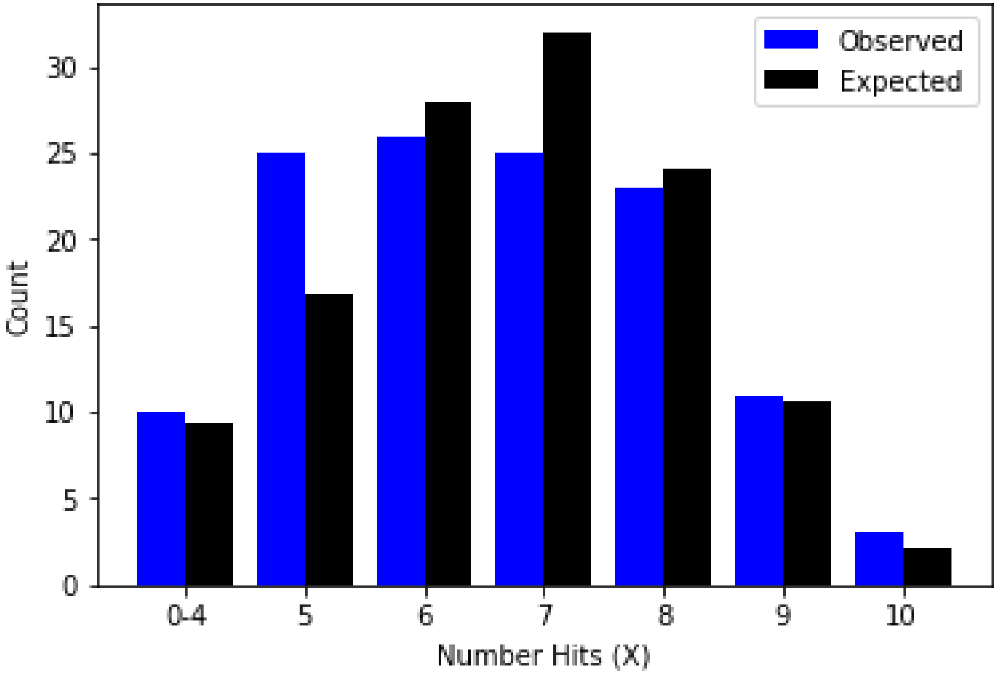
x_i	0-4	5	6	7	8	9	10
E_i	9.4	16.8	28.0	32.0	24.0	10.7	2.1

Chi-squared (χ^2) Test Statistic

$$T = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

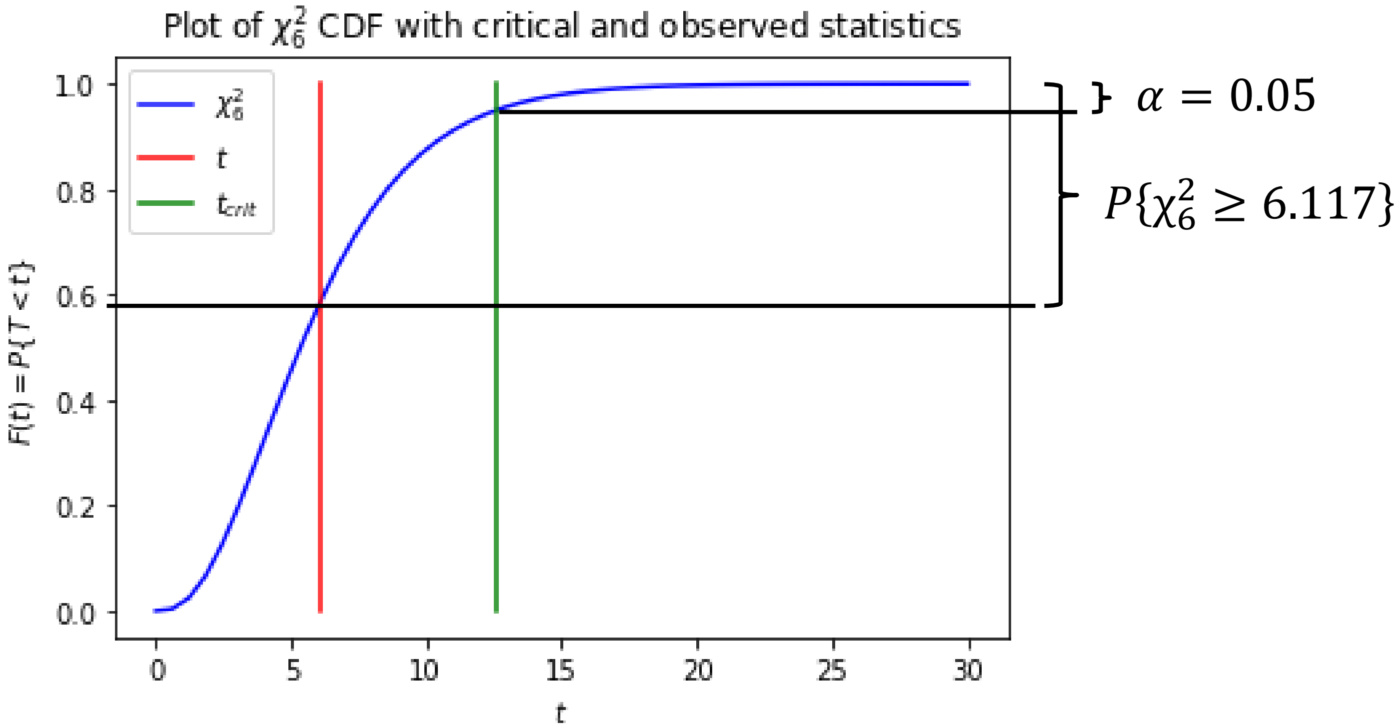
- N samples of n discrete values
- O_i observed samples for each value i
- E_i expected samples for each value i, $E_i = N * p(x_i)$
- p−value $= 1 - F_{\chi_k^2}(T)$
  - $F_{\chi_k^2}$: chi-squared CDF with $k = n - 1 - c$ degrees of freedom
  - c: number of distribution parameters estimated from samples
- Reject null hypothesis if p−value $< \alpha$ (bad distribution!)

Dice Roller Statistic



x_i	O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
0-4	10	9.4	0.036
5	25	16.8	4.005
6	26	28.0	0.142
7	25	32.0	1.529
8	23	24.0	0.041
9	11	10.7	0.011
10	3	2.1	0.352
		$T =$	6.117
		$k =$	7-1=6

Dice Roller Test



$$p\text{−value} = 1 - F_{\chi^2_6}(6.117) = 0.41$$

→ Cannot Reject H_0: data similar to binomial distribution

→ 41% chance of a more extreme statistic under H_0

Goodness of Fit Tests for Continuous Data

Continuous Goodness of Fit Tests

A **Goodness of Fit** test checks if samples observed of a random variable come from a particular distribution

- Validate process generators with commonly-used distributions (e.g. Normal, Exponential)
- **Chi-squared (χ^2) Test**: group continuous variables in bins (same as discrete case)
- **Kolmogorov-Smirnov Test:** compare observed CDF (discrete) with expected CDF (continuous)
- **Anderson-Darling Test:** variation of K-S test

Kolmogorov-Smirnov (K-S) Test

- **Kolmogorov-Smirnov (K-S) test** evaluates whether a set of samples follow a hypothesized *continuous* distribution

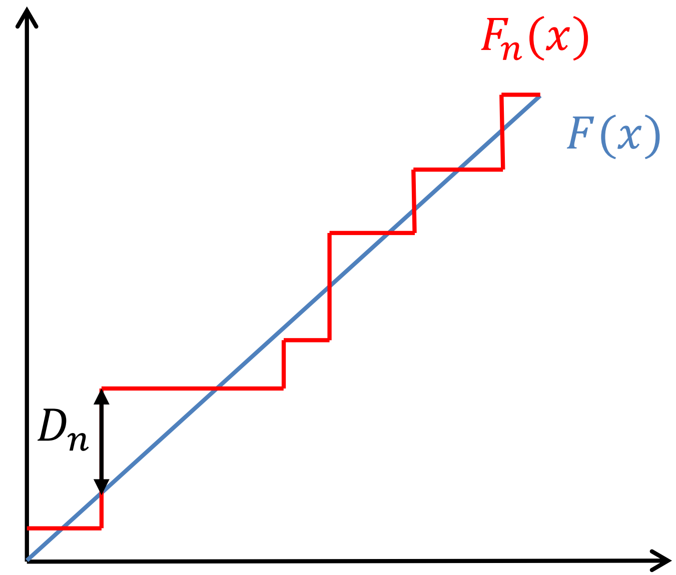
  H_0: the data are consistent with a specified distribution

  H_a: the data are not consistent with a specified distribution

- Compares the following:
  - *Expected* or *theoretical* CDF
  - *Observed* or *actual* CDF

K-S Test Statistic

- Test statistic D_n considers:
  - Largest vertical distance between observed and expected CDF
  - $F_n(x) = \frac{\text{\# samples} \leq \text{x}}{\text{\# samples}}$
  - $D_n = \max_{1 \leq i \leq N} \left(F(x_i) - F_n(x_{i-1}), F_n(x_i) - F(x_i) \right)$

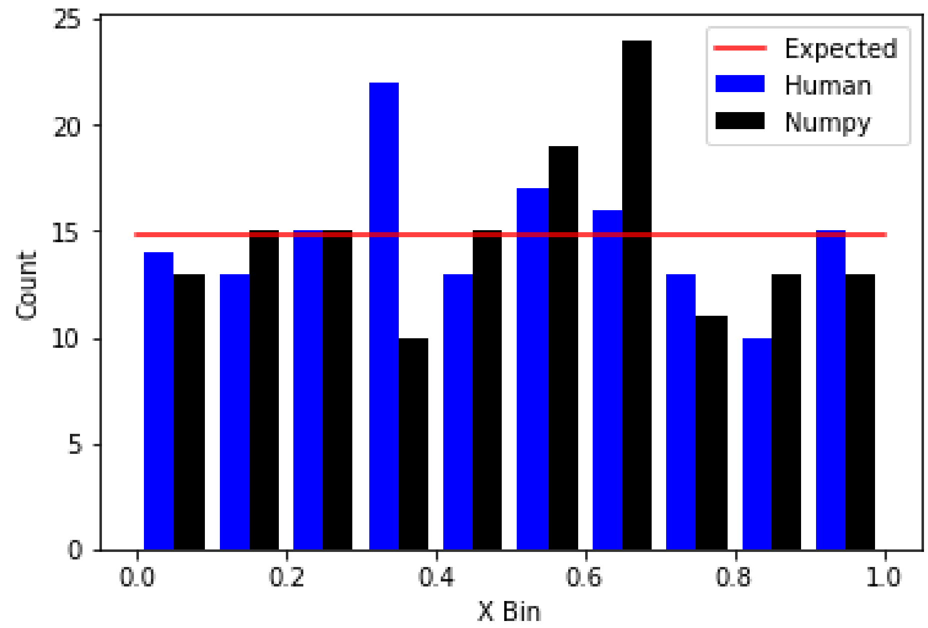


- Reject null hypothesis at significance level α if:

$P\{D_n \geq d_n\} = p{-}\text{value} < \alpha$ (use statistical software)

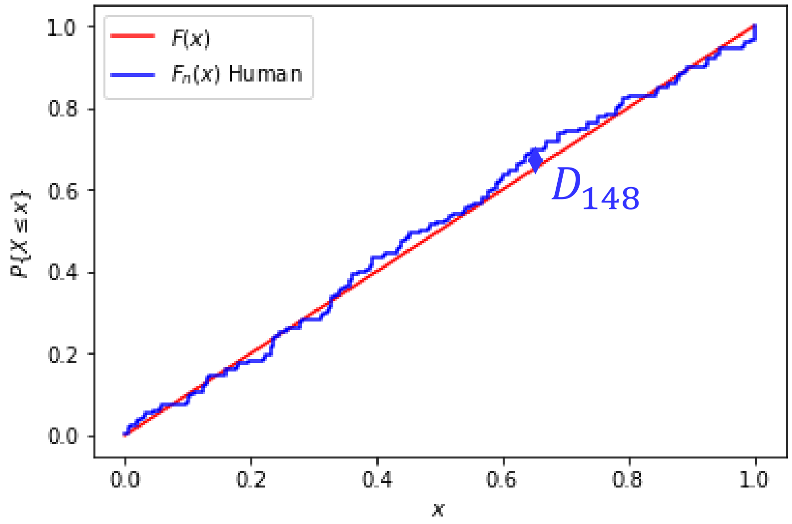
Human RNG Data

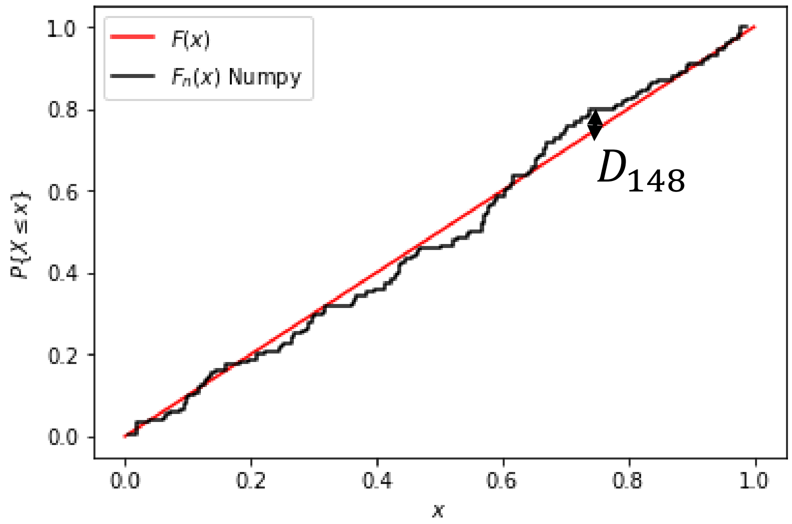
- Two U(0,1) sources with 148 samples:
  - Human samples
  - Numpy random.rand()
- Compare both with expected uniform (0,1) distribution
  - Examples use **`scipy.stats.kstest`**



Human RNG Statistic/Test

- Human samples
  - $D_{148} = 0.051$
  - $p-\text{value} = 0.84$
- Numpy samples:
  - $D_{148} = 0.067$
  - $p-\text{value} = 0.52$
- Both human and Numpy samples are similar to uniform (0,1) distributions!





Based on Hazelrigg (2003)

Engineering Model Validation

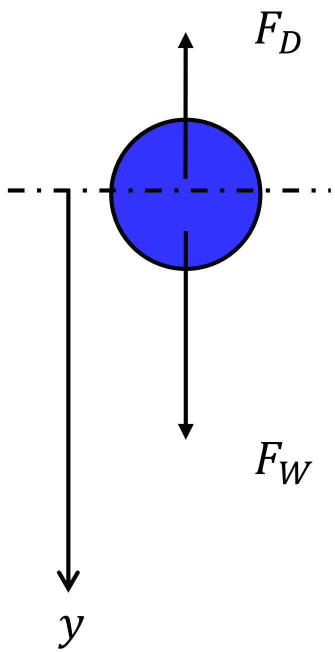
Scientific vs. Engineering

Based on Hazelrigg (2003)

- Science uses **models** to express causality
  - *Laws of nature* are “good” models that appear to be invariant and fixed over time and space
  - For example: $F = ma$
- Engineering uses **models** to represent physical artifacts and their behaviors
  - Necessary simplifications of reality, never perfect
  - Errors are dependent on the context

Example: Falling Object

Based on Hazelrigg (2003)



Natural Law

$$y = y_0 + \iint_{t_0}^{t} \frac{\sum F}{m} dt$$

Engineering Model

$$F_D = 0$$

$$F_W = mg$$

- No drag
- Earth surface

$$F_D = \frac{1}{2}\rho v^2 C_D A$$

$$F_W = G\frac{m_1 m_2}{r^2}$$

- Form drag
- Two-body gravity

Validity of engineering model is completely dependent on the application context

Use of Engineering Models

Based on Hazelrigg (2003)

- Not all engineering models must give **absolute measures** to be useful
- Models to inform decisions only need to **discriminate** between alternatives
  - Which is the better design to reduce drag?

vs.

- **Rational** decisions select the best alternative

Design Decisions

Based on Hazelrigg (2003)

- Must consider three things for design decisions:
  - **Alternatives**: various courses of action available
  - **Perceptions**: expectations of alternatives
  - **Preferences**: desirability of outcomes
- Only individuals make decisions
  - Act on behalf of own perceptions and preferences
  - Commit resources in the present
  - Seek preferred outcomes in the future

Models as Information Sources

Based on Hazelrigg (2003)

- **Model information quality** describes the probability a preferred choice leads to the most desirable outcome

- A model with **perfect information** guarantees the preferred choice has the best outcome
- **Valid models** produce high-quality information that leads to preferred choices/outcomes

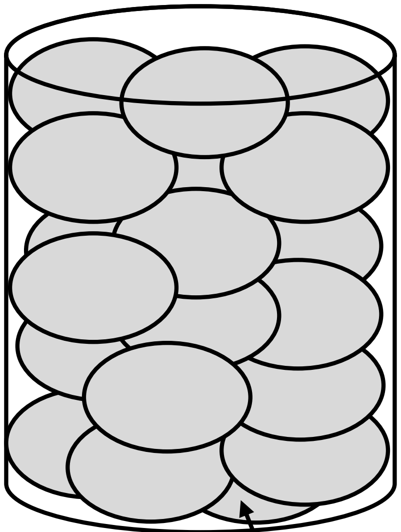
Guessing Game

Based on Hazelrigg (2003)

- **How many M&Ms are in the jar?**
- One guess per person, closest *without going over* wins the jar
- How can models provide information for this decision?
  - What are the alternatives?
  - What are your perceptions?
  - What are your preferences?

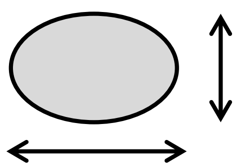
M&M Model

Based on Hazelrigg (2003)



$V = 1890 \pm 2\%\ \mathrm{mL}$ $\quad\mu = 0.55 \pm 20\%$

$V_c = \dfrac{\pi}{6} d^2 t$



$t = 0.6 \pm 10\%\ \mathrm{cm}$

$d = 1.4 \pm 10\%\ \mathrm{cm}$

$$N = \frac{V}{V_c}\mu = \frac{V}{\frac{\pi}{6} d^2 t}\mu$$

Volume of container:

$V \sim \mathrm{triangular}(1852, 1890, 1930)$

Diameter/thickness of M&M:

$d \sim \mathrm{triangular}(1.26, 1.4, 1.54)$

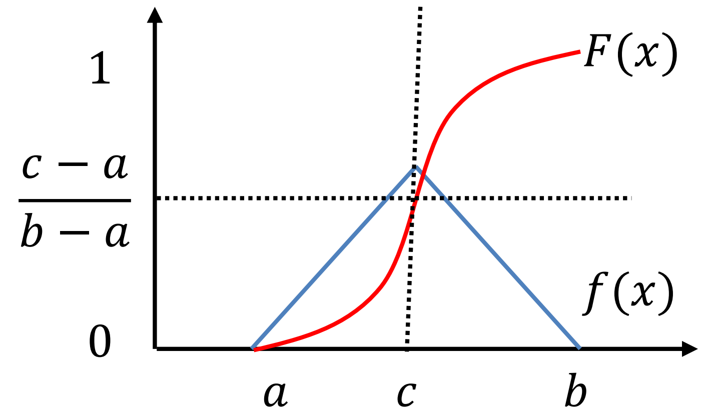
$t \sim \mathrm{triangular}(0.54, 0.6, 0.66)$

Packing factor:

$\mu \sim \mathrm{triangular}(0.44, 0.55, 0.66)$

Triangular Process Generators

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & a \le x \le c \\ \frac{2(b-x)}{(b-a)(b-c)} & c < x \le b \end{cases}$$



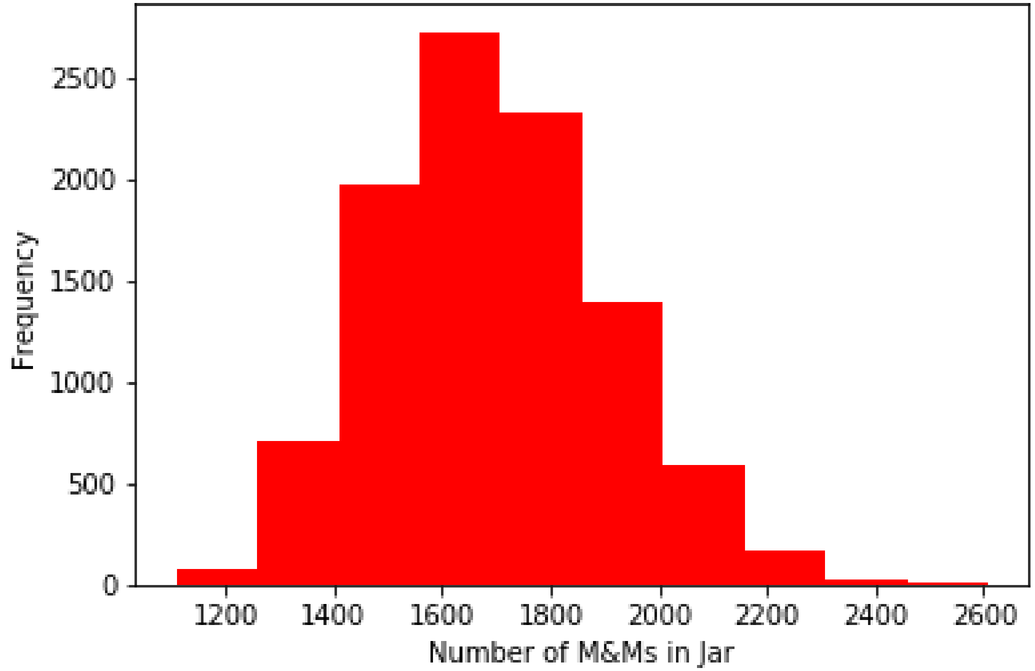
$$F(x) = \begin{cases} \frac{(x-a)^2}{(b-a)(c-a)} & a \le x \le c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & c < x \le b \end{cases}$$

$$x = F^{-1}(r) = \begin{cases} \sqrt{r(b-a)(c-a)} + a & r \le \frac{c-a}{b-a} \\ b - \sqrt{(1-r)(b-a)(b-c)} & r > \frac{c-a}{b-a} \end{cases}$$

(even easier in Python… `np.random.triangular(a,c,b)`)

Derived Distribution of N

- 10000 samples
- Mean: 1697
- Std. dev: 210.8
- Std. error: 2.1



Modeling Preference

- Derived state variable w shows whether a choice x wins the jar of M&Ms
  - Determined based on true number $N^\star$
  - Compare versus all others' choices $\boldsymbol{y}$

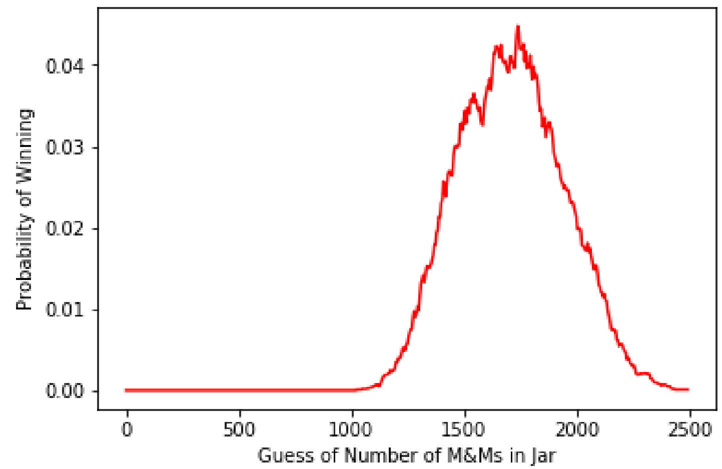
$$w(x, \boldsymbol{y}) = \begin{cases} 1 & if\ x \leq N^\star\ and\ x \geq j\ \forall\ j \in \boldsymbol{y}\text{: } j \leq N^\star \\ 0 & \text{otherwise} \end{cases}$$

- Model should estimate $p(w(x))$ from many trials
  - Find x to maximize probability of winning!
  - Try for 50 opponents…

Monte Carlo Simulation

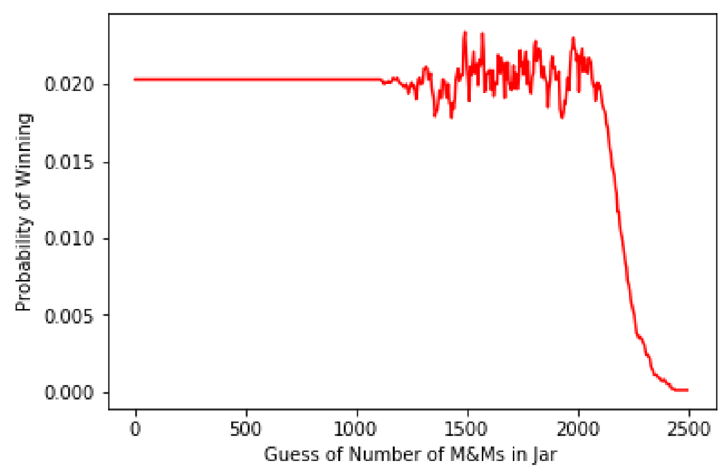
- Simple expectation of others' choices:

$$y \sim \text{triangular}(500{,}1600{,}2500)$$



- More advanced expectation of others' choices:

$$y \sim N$$



Committing Resources

- For question 7.1 on this week's assignment, take some time to do your own analysis…
  - "Le Parfait Super Terrines" 500ml jar, **overfilled** above max
  - Standard milk chocolate M&Ms
- Optional: submit your **one** final choice here:

goo.gl/ueFSFF