Statistical Methods for Insurance: Statistical distributions

Di Cook & Souhaib Ben Taieb, Econometrics and Business Statistics, Monash University
W3.C1

Overview of this class

- Quiz 1 solution
- Random numbers
- Mapping random numbers to events for simulation
- Statistical distributions
- READING: CT6, Section 1.3-1.9

Random numbers

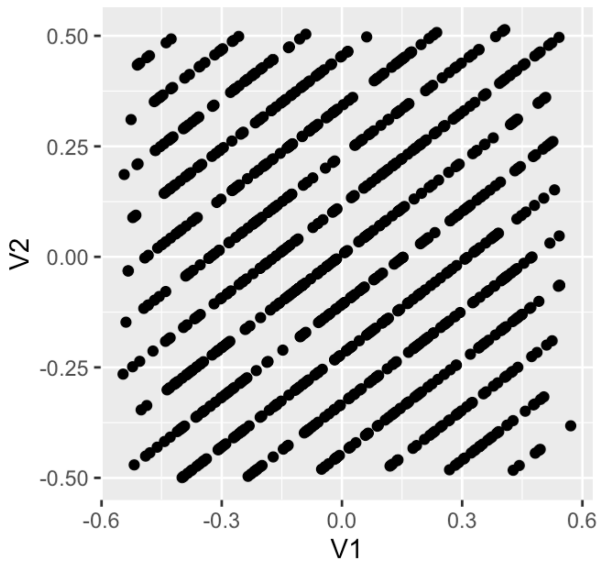
- True random number generators: Radioactive decay, electromagnetic field of a vacuum
- Computers only technically provide pseudo-random numbers, using deterministic process, e.g linear congruential, for large a, b, m

$$X_{n+1} = (aX_n + b) \;\; mod \;\; m$$

RANDU - a bad PRNG

- Used in the 60s and onwards

$$X_{n+1} = 65539 X_n \ \ mod \ \ 2^{31}$$



Mersenne Twister

- algorithm is a twisted generalised feedback shift register (TGFSR)
- based on a Marsenne prime, $2^m - 1$
- most commonly used today
- each integer will occur the same number of times in a period

Using random numbers

- Random number tables deliver single digits 0, 1, …, 9
- When using these you need to ensure that you map these digits or combinations of the digits to match the probabilities of events
- For example, use random numbers to sample students from class
  - There are 105 students in the class
  - Need to use three sequential digits
  - BUT there are 1000 three digit numbers, so either we will throw away 895 of them, or we could map a person to multiple numbers (9) and throw away only 55
  - If any person is selected more than once, throw out repeats

Estimate the proportion of 2420:5242

Class list:

First	Last	numbers
Reece	Agiazis	001,002,003,004,005,006,007,008,009
Yuan	An	010,011,012,013,014,015,016,017,018,019
Eric	Au	020,021,022,023,024,025,026,027,028,029
Travis	Barr	030,031,032,033,034,035,036,037,038,039
Prangana	Barua	040,041,042,043,044,045,046,047,048,049

Set of random digits

```
#>   [1] 9 1 6 9 5 6 7 2 8 7 9 3 1 8 8 5 6 6 9 7 8 8 8 8 1 8 5 7 4 0 6 2 8 3 2
#>  [36] 3 6 2 3 2 0 2 5 9 4 7 2 9 9 2 5 3 8 2 0 9 8 9 0 3 3 4 3 3 0 6 5 6 7 5
#>  [71] 1 6 8 1 4 8 8 5 4 9 3 9 1 1 6 4 5 6 5 8 4 7 9 2 5 7 4 4 0 8 6 9 6 8 1
#> [106] 3 9 7 4 9 6 3 1 8 5 0 5 5 6 9 0 6 3 6 8 3 0 9 9 5 3 7 9 8 4 0 7 3 9 4
#> [141] 8 5 2 0 7 7 2 7 8 8 8 1 3 8 3 3 8 3 3 9 2 8 5 2 2 7 6 9 2 7 9 6 2 9 6
#> [176] 7 7 8 8 6 2 0 5 9 6 5 1 9 7 3 3 1 5 7 0 1 0 9 4 0 8 6 7 7 2 4 7 0 2 4
#> [211] 7 5 3 6 1 7 5 5 2 6 1 4 0 2 1 2 1 6 9 9 0 9 8 0 9 8 4 0 9 8 2 4 0 2 3
#> [246] 8 8 8 0 2 7 9 9 4 8 8 4 2 9 1 0 0 4 5 2 4 1 2 7 9 9 9 6 3 5 1 0 3 1 8
#> [281] 4 6 7 9 9 6 6 5 2 4 1 5 6 5 7 2 1 1 7 7 1 7 9 0 1 5 1 7 8 7 3 2 7 4 2
```

Just do it …

True proportion of 2420:5242 is 0.59. Sample 20 students and check this.

Simpler approach computationally

- Use a sample function

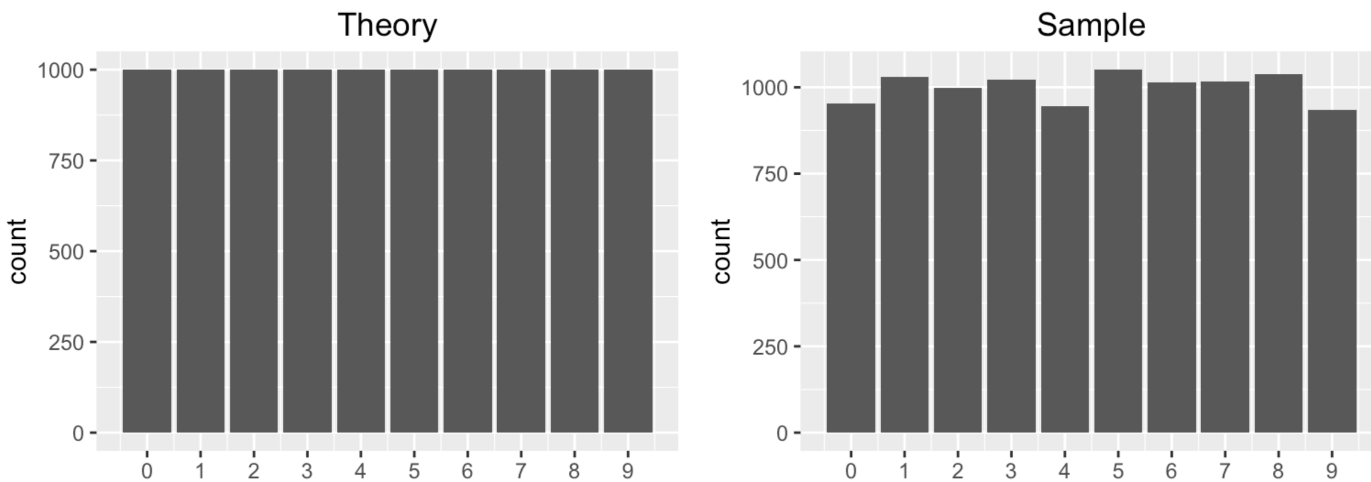
```
sample_class %>% sample_n(20)
```

First	Last
Carien	Leushuis
Ken	Tan
Yi	Chin
Chee	Lau
Benjamin	Mok
Jun	Poon
Menglei	Cui
Jenipher	Maloba
Joshua	Prameswara
Guizhen	Li

Statistical distributions

- Uniform
- Normal
- Exponential
- Binomial
- Pareto
- Weibull
- Gamma
- Lognormal

Random numbers = Uniform



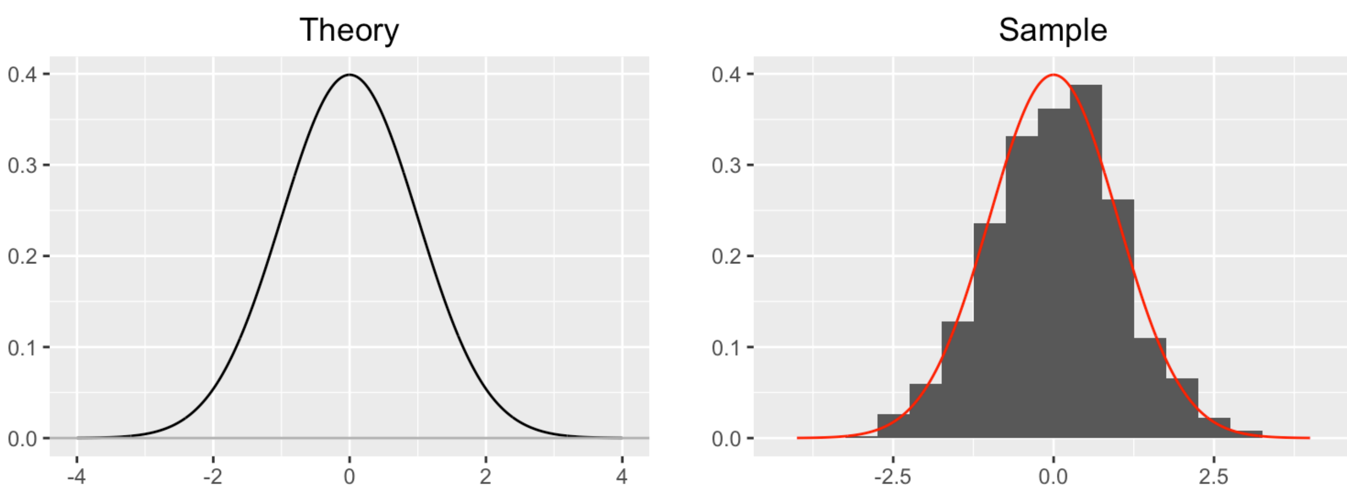
- symmetric, unimodal, uniform
- e.g. $U\{0, \ldots, 9\}$
- e.g. $P(X = x) = f(x) = 1/10, \quad x \in \{0, \ldots, 9\}$

13/26

Normal distribution

- Gaussian, bell-shaped
- symmetric, unimodal
- $N(\mu, \sigma)$

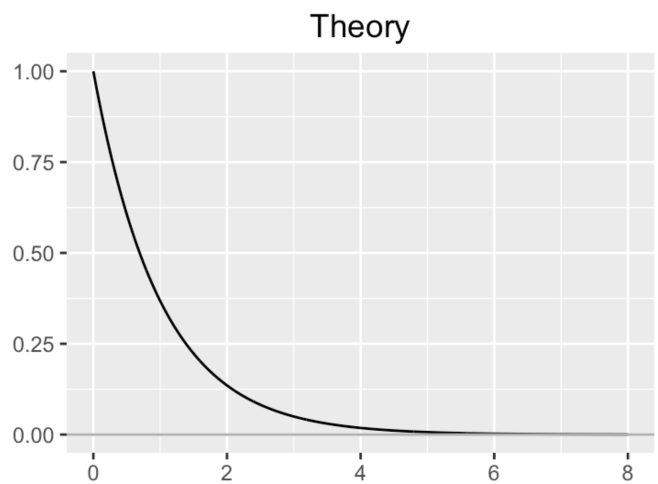
$$f(x \mid \mu, \sigma) = \frac{1}{\sqrt{2}\pi\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

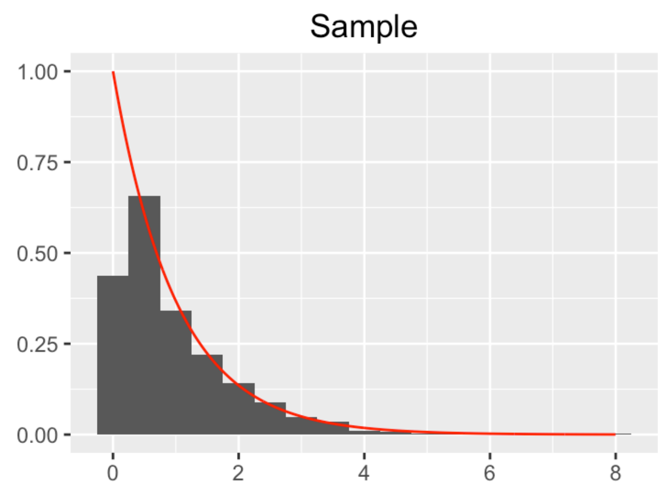


Exponential distribution

$$f(x \mid \lambda) = e^{-\lambda x} \quad x \geq 0$$

- right-skewed, unimodal
- $Exp(\lambda)$
- Arises in time between or duration of events, e.g. time between successive failures of a machine, duration of a phone call to a help center

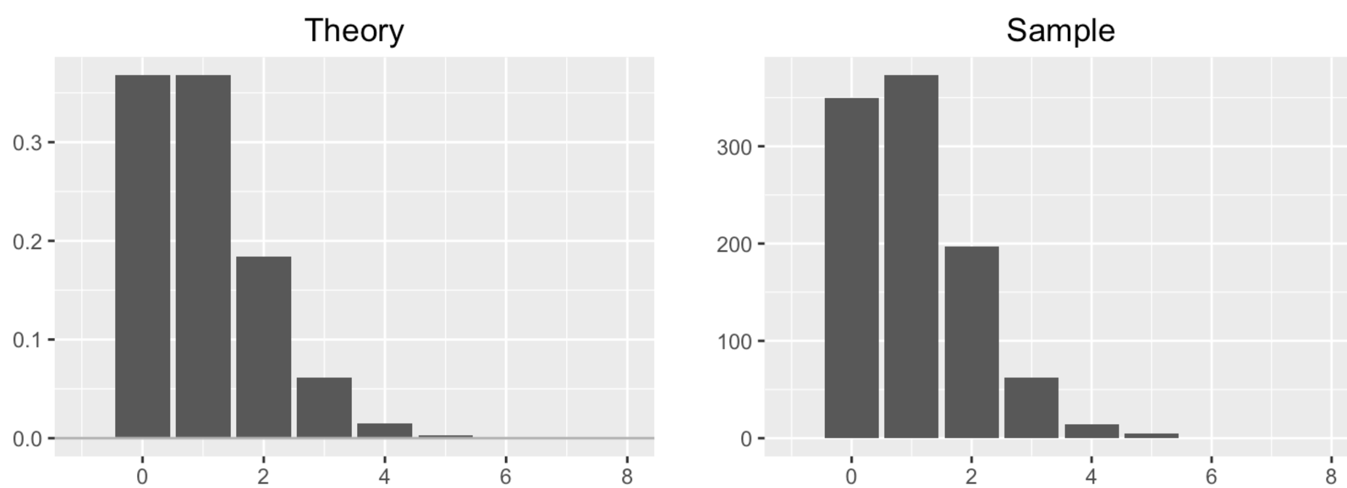




Poisson distribution

$$P(X = x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x \in \{0, 1, 2, \dots\}$$

- discrete, right-skewed, unimodal
- Arises when counting number of times event occurs in an interval of time, e.g. the number of patients arriving in an emergency room between 11 and 12 pm

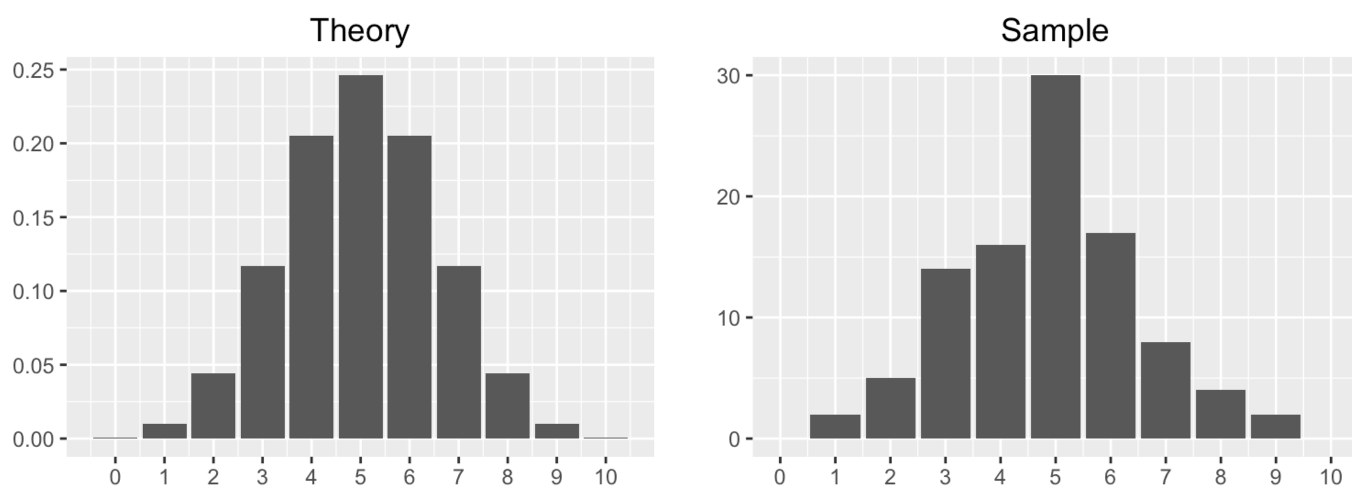


16/26

Binomial

$$P(X = x \mid n, p) = \binom{n}{p} p^x (1-p)^{n-x} \quad x \in \{0, 1, 2, \ldots, n\}$$

- discrete, unimodal, right- or left-skewed or unimodal depending on p
- Arises from counting the number of successes from n independent Bernouilli trials, e.g. the number of heads in 10 coin flips

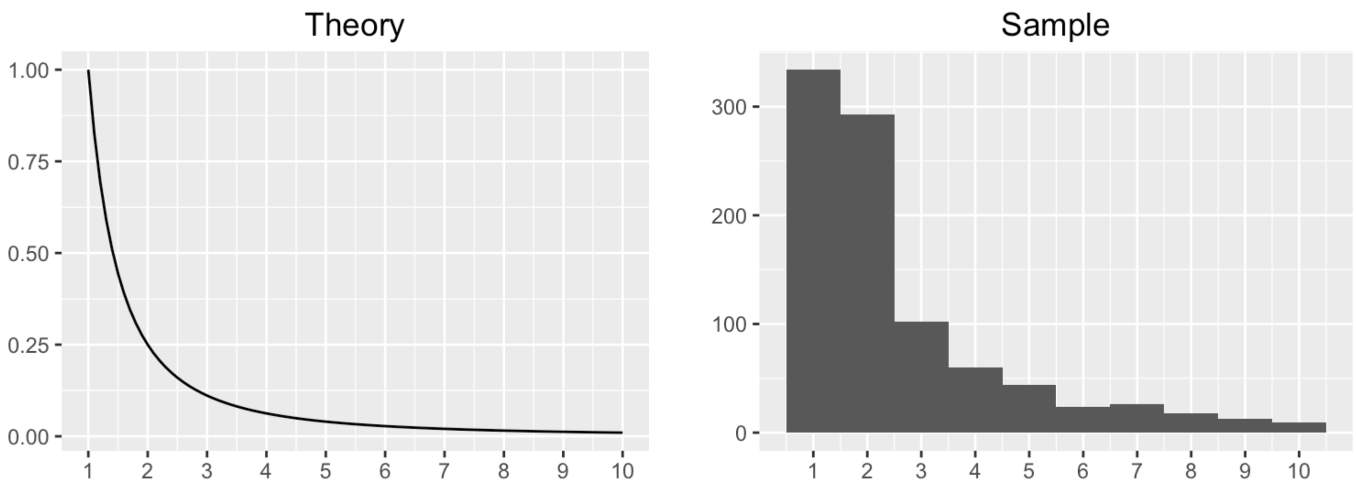


17/26

Pareto

$$f(x \mid \alpha, \lambda) = \frac{\alpha\lambda^{\alpha}}{(\lambda + x)^{\alpha+1}} \quad x > 0, \alpha > 0, \lambda > 0$$

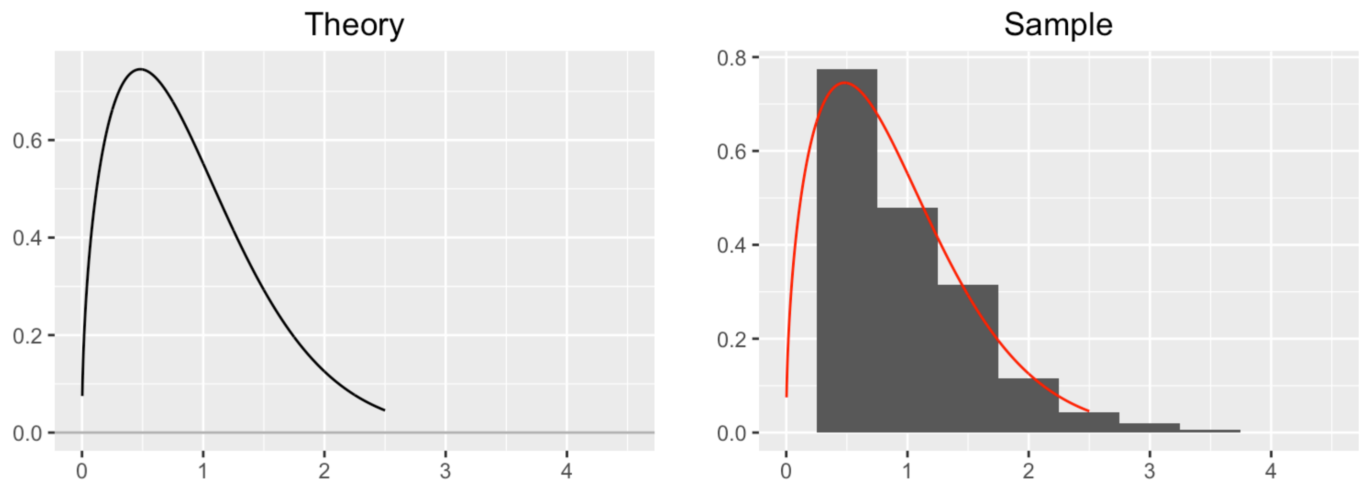
- Used to describe allocation of wealth, sizes of human settlement
- Heavier tailed than exponential distribution



Weibull

$$f(x \mid \lambda, k) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{(-x/\lambda)^k}, \quad x \geq 0$$

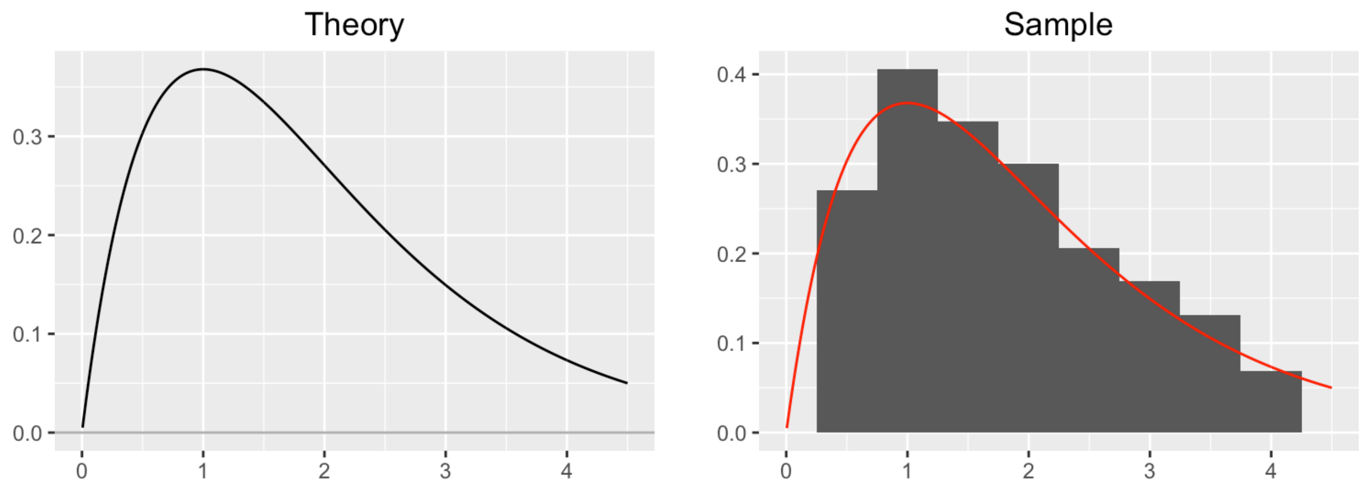
- used for particle size distribution, failure analysis, delivery time, extreme value theory
- shape changes considerably with different k



Gamma

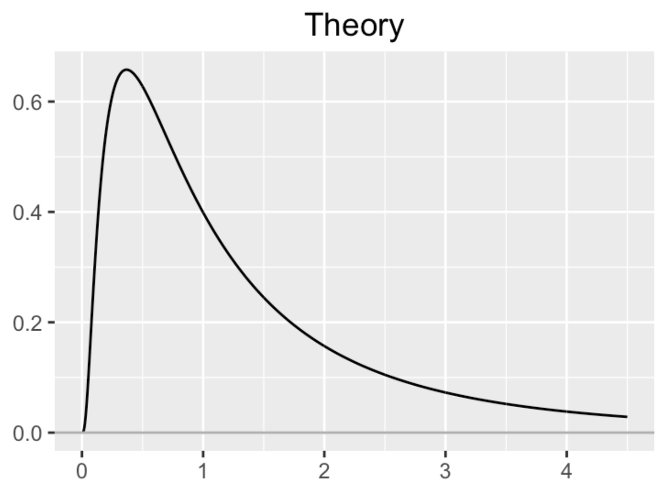
$$f(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-x\beta}, \quad x \geq 0 \;\; \alpha, \beta > 0$$

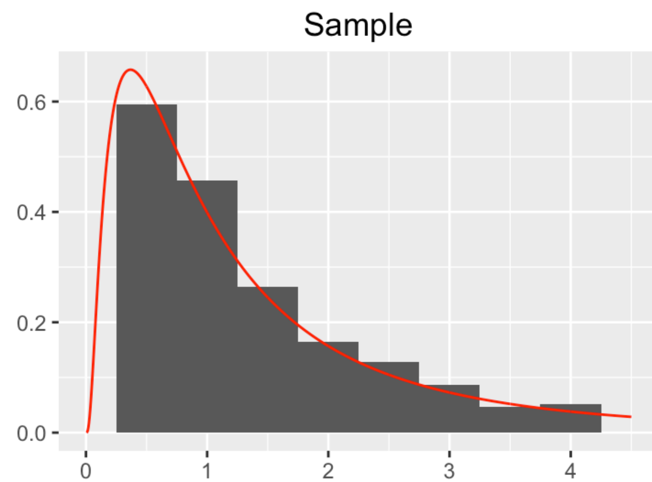
- Generalisation of exponential distribution, and also χ^2
- α changes shape substantially
- used to model size of insurance claims, rainfall



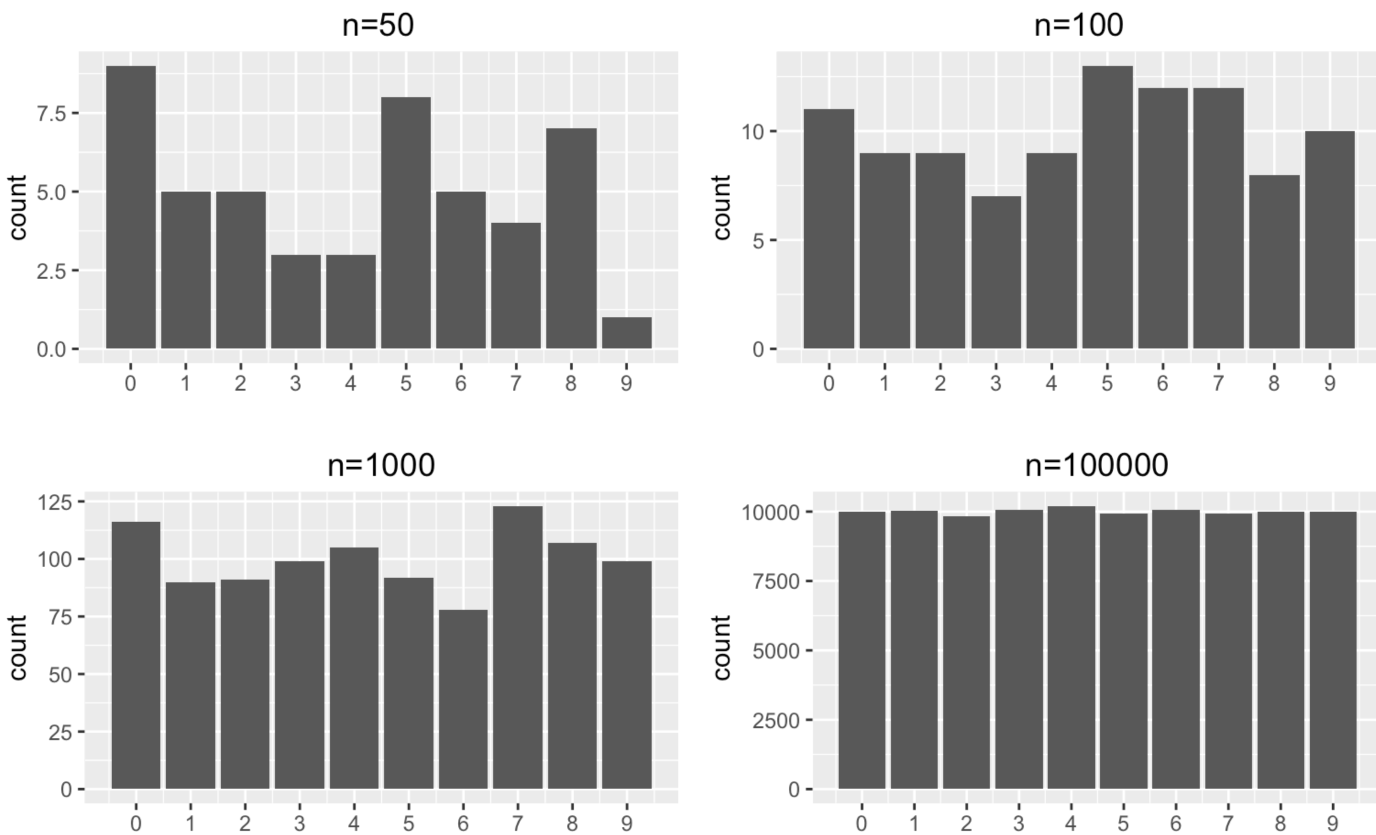
Lognormal

- Also called Galton's distribution
- Generated when $Y \sim N(\mu, \sigma)$, and study $X = exp(Y)$
- used for modeling length of comments posted in internet discussion forums, users' dwell time on the online articles, size of living tissue, highly communicable epidemics





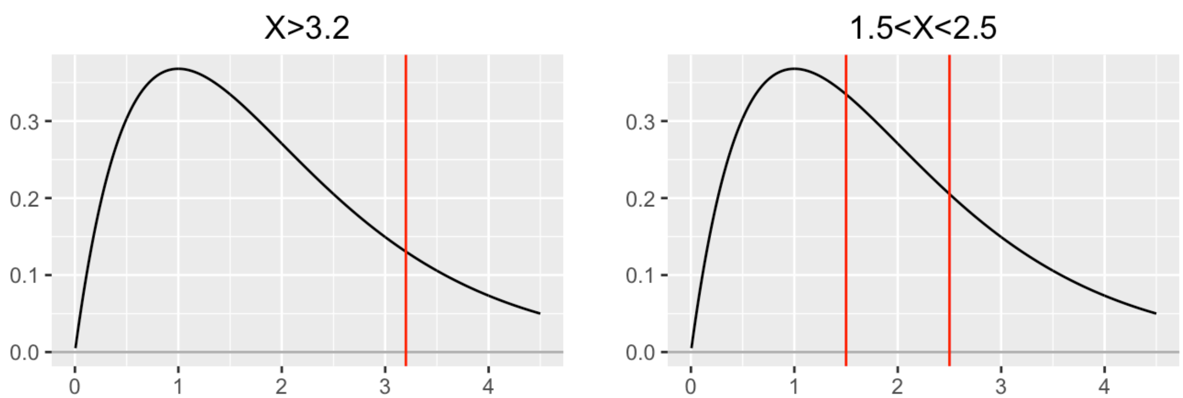
Sampling variability



Probability calculations

- Probability density functions are useful for computing expected quantities
- E.g. Gamma(2,1), what is the probability of seeing $X > 3.2$, or $1.5 < X < 2.5$

```
pgamma(3.2, 2, lower.tail=FALSE)
#> [1] 0.17
pgamma(2.5, 2) - pgamma(1.5, 2)
#> [1] 0.27
```



Your turn

- Continuous distributions: Area under the curve = ______
- Discrete distributions: Sum of probabilities = ______

Resources

- NIST Statistics Handbook
- random.org
- Radioactive decay
- electromagnetic field of a vacuum
- wikipedia

Share and share alike